

Academy elections in a muddle

The Soviet Academy of Sciences has been made a laughing-stock by the muddle over its constituency election for the Supreme Soviet. But the difficulty should have been anticipated.

THE continuing pantomime (see page 364) of the failure of the Soviet Academy of Sciences to elect a score of people as its representatives in the new Chamber of the People's Deputies is neither a surprise nor a criticism of the academy. On the contrary, it may more properly be regarded as a sign of grace. The academy's failure even to nominate enough people to fill the 25 places it had been offered in the new chamber (in the end, there were only 23 nominees, so that five places were relinquished) is a measure of the diversity of opinion within the academy, and of its members' willingness to behave as they believe. Is that not a virtue, rather than the opposite?

Even so, the conduct of the election, and its outcome, will not enhance the academy's reputation within the Soviet Union or elsewhere. Nor will it do much for the reputation of the Soviet electoral system, hastily devised last year as part of the cutting edge of *perestroika*. There are too many wheels within wheels for anybody's comfort. One difficulty is that there are too many kinds of constituencies. Some are geographical, some geographical constituencies include groups of others, while other constituencies are professional (such as the academy) or defined by other special interests (such as the company of Soviet philatelists). Another difficulty is that the rules by which professional and other societies have approached the process of election have been vague, to say the least.

Within the academy, the first step (in January) was a kind of popularity contest in which individual institutes put forward names for nomination. Sakharov and Sagdeev (in that order) came out on top, but then failed to win half the votes that might have been cast at the special conference a few weeks later. Now, 15 out of the 23 nominees still in the race have fallen for the same reason, no doubt victims of the advertised determination of those who wished to vote for Sakharov and Sagdeev to vote against all other candidates. Why, despite all the secret ballots, should candidates overwhelmingly popular at one stage of the election fail to be elected at the next? The immediate explanation is that the electorate changed (and narrowed) between the beginning of January and the end. And the explanation of that is that the academy, while seeking to give democratic processes an airing, did not dare go far down the road of one man, one vote.

Yet the Chamber of Deputies being elected is an important part of a new and more democratic system. Although it will meet only infrequently, it will have the crucial task of choosing the new Supreme Soviet, intended to become a full-time legislature under Mr Mikhail Gorbachev himself, who will be its president. It will have influence, even power. All the more reason why it should seem legitimate. Perhaps its first task should be to revise the law under which it is being elected, in the process abolishing the professional constituencies and the underlying principle of "most men, one vote: some men, two votes".

For the Soviet academy, the lessons to be drawn from this affair go even deeper. In the Soviet Union now, elections for institute directors, laboratory heads and others are all the rage. But on a matter such as the election of people to the Chamber of Deputies, which has nothing to do with the management of the academy's own over-complicated business, it is absurd that the academy should have canvassed the wishes of those working in its institutes and then have allowed them to be denied. If laboratory workers want Sakharov and Sagdeev to represent them, why should they not have their way? □

Cold (con)fusion

Reports that an account of cold nuclear fusion is soon to appear in this journal are premature.

INCREASINGLY alert coverage of science by newspapers, taken as a sign of increasing public interest, is something that readers of this journal should applaud. But when scientists find themselves reading about their colleagues' discoveries in newspaper columns before anything has been submitted, let alone accepted or published, in a research journal, there is cause to be worried. No one was more surprised than the editors of *Nature* to learn, on reading last Friday's *Wall Street Journal* that two papers on room-temperature nuclear fusion (see page 364) would probably appear simultaneously in this journal, perhaps in May.

That editors of scientific journals are annoyed by this kind of event is due neither to sour grapes nor to a cabalistic devotion to secrecy. The procedure of peer-



review, slow and irksome as it sometimes can be, has evolved to protect not only journals but scientists and science itself from a barrage of unsubstantiated claims which later have to be withdrawn. The more outlandish the claim, the less likely authors seem willing to submit to careful scrutiny by knowledgeable people working in related areas. There are understandable reasons for this reticence. It is the rare piece of research indeed that both flies in the face of accepted wisdom and is so compellingly correct that its significance is instantly recognized. Authors with unique or bizarre approaches to problems may feel that they have no true peers who can evaluate their work, and feel more conventional reviewer's mouths are already shaping the word "no" before they give the paper much attention. But it must also be said that most unbelievable claims turn out to be just that, and reviewers can be forgiven for perceiving this quickly.

But unsubstantiated claims often make interesting news, which journalists have a duty to report even if they do so sceptically. Unlike other 'news' stories that appear in daily newspapers, 'science news' stories have no clear date when they become news worth printing. Is a scientific discovery news from the day it is made, the day it is submitted to a journal, the day the journal accepts it, or the day it is published? Reasonable people will differ on which is the correct answer.

Even with the best of luck and planning, this journal cannot take a scientific paper from receipt to publication in less than about a month. Last week's issue contains two examples, concerning the new pulsar in supernova 1987A, in which the researchers concerned were able to keep quiet for the required few weeks. But there are those who feel compelled, either for money or fame, to ballyhoo discoveries before anyone can reasonably judge their merits. Superconductors, AIDS treatments and new energy sources bring behind them patent lawyers and venture capitalists, to whom breaking the press embargo of an academic journal probably seems a negligible misdemeanour.

But patents that turn out to be worthless and investments that disappear may demonstrate the value of cautious peer-review to those who now think of it as a fusty institution much loved by pedants. The sceptical tone adopted in many newspaper accounts of cold fusion, and the emphasis given to the fact that no paper has yet been accepted for publication in any scientific journal, indicates that journalists (who learn nearly as fast as they write) appreciate the pitfalls. Many learned a lesson two years ago, when reports of superconductivity at increasingly incredible temperatures were appearing daily. It did not take long for science writers to realize something was amiss, and after seeking professional advice they began asking for specific evidence of superconductivity, such as magnetic-flux expulsion, in addition to claims of sudden drops in resistance. An obvious need existing, peer-review was spontaneously re-invented. □

Words by machine

Computerizing the *OED* should help its publishers to remedy the dictionary's most obvious defects.

STEPHEN J. Gould is right to welcome the new edition of the *Oxford English Dictionary* as a landmark in the intellectual life of the English-speaking world (see page 385). It may be even more than that. That the dictionary of a language should be constructed on historical principles, recording all the usages of all words without the disparagement of any, was in itself a radical notion even further out of tune with the mood of Victorian England when the dictionary was begun than with the still-prescriptive inclinations of our times. The result, as Oxford University Press is eager to proclaim, is not so much a dictionary as a living record of the changing English language, or at least of that part of it used by the British. Will the appearance of the new edition of the dictionary spark the imitation by other-tongued linguistic chauvinists that could further augment its value by providing comparative records of how languages evolve?

For the British (and for the dictionary's publishers), several questions arise. The dictionary now published as ink on paper is not the real thing, but a first approximation to it. Words and meanings are accumulating all the time. The publishers plan to put out another version in a few years on compact disks readable by laser. No doubt successive versions will be numbered much as software manufacturers sell new versions of their products, in a decimal notation. Because the dictionary is built on historical principles, that should not be as confusing as it seems: new versions will include their predecessors. But, now that the Oxford University Press has survived its first serious brush with computer technology, there is also a case for making the dictionary accessible on-line. There is also now an opportunity to use the new technology for capturing examples of the use of new words, or of old words in new meanings, whose compilation has been the back-breaking and sometimes disheartening part of the labours of the past century. It is not merely that the publishers will thereby be saved time and money, but that they will then be less likely to neglect the maintenance of a important scholarly resource than they were in the 1930s, when work on the dictionary was halted for a quarter of a century.

That is also how best to make progress towards remedying even the new edition's most obvious deficiencies, of which there are two. First, the dictionary's treatment of even common usages in science is hesitant at best, as may be told from the article on "anti-particle" in the new edition. Second, it is too dependent on British sources. When everything had to be done by hand, there may have been no choice. Now there is a chance to make it into an English dictionary in the full sense. That should not be neglected. □

Shake-up in store for Britain's biologists

- New organization eliminates fragmentation
- Single funding source proposed for all biology

London

THE biological sciences in Britain's universities are underfunded, short of manpower, and teaching and research are "undesirably fragmented", according to a committee set up by the Universities Grants Committee (UGC) to advise on how academic biology could be 'rationalized'. The results of the review have been shrouded in secrecy until now, and there is likely to be heated debate about the conclusions at a meeting this week of the heads of Britain's biology departments.

Similar reviews of chemistry and physics last year concluded that to maintain quality with the limited funds available resources should be concentrated in large departments—those with at least 20 full-time staff. But the biology committee concludes that resources for biology must expand. Professor Richard Southwood of the University of Oxford, who chaired the UGC committee, says biology will be the "leading science of the twenty-first century". He complains that biology has always been underfunded because when interest in the subject increased in the 1970s, a period of university expansion had ended and there were increasing constraints on growth. An injection of funds is now necessary to compensate for this, he says.

The committee says that the main problem facing biology is the fragmentation of teaching, research and of the funding mechanisms. To eliminate fragmentation in support, the report recommends money for biology be provided from one research council instead of the current four. Most support for biology is now channelled through the Science and Engineering Council (SERC). But to avoid competition with the expensive demands of the physical sciences and engineering, the responsibility for biology should not be vested in SERC but in a new biological sciences council.

For research and teaching, the report recommends that all biology departments be grouped into a single organization, within which there should be two distinct subgroups. One, termed M, would focus on molecular sciences, including biophysics, biotechnology, physical biochemistry and molecular genetics. The other, called B, would focus on traditional areas of biology such as physiology, ecology, entomology and evolutionary studies. Most biologists do not dispute that more integration of their departments is desirable, but the M/B

idea is nonetheless controversial. To ensure that all facets of biology are covered in each department, the committee says there would need to be 5 research groups each with at least 4 members of staff, so each department would contain at least 20 members of staff.

Meeting the requirements of size and covering the defined areas at the same time would require "an enormous amount of restructuring" in universities, said Professor Christopher Arme of the University of Keele, though he agrees that larger departments than exist at present would be desirable. Others have a different view. Professor David Cove of the University of Leeds argues that small departments can be more efficient than large ones, and a greater amount of personal attention to students is rewarded with a better quality of student.

Southwood stresses that closure of any biology department would be short-sighted. The predicted demographic dip in the 1990s leading to fewer young people of university age is a misleading indicator of future demand. There is likely to be an acute shortage of suitable personnel for posts in biology in universities and research-council institutes in the future, says the report. It supports the schemes of the UGC and the Royal Society to tackle this problem, but warns against central planning in relation to subject and place of study.

One of the concerns of the committee is that traditional areas of biology are increasingly being neglected, and it believes that there should be an equal number of M and B departments. Undergraduates would take a three-year degree course in one department, but should have a basic understanding of the biology studied by the other groups. The committee also warns against central planning in distribution of research funds. It is impossible for committees to identify topics that will lead to significant advances in the future, and best value for money will be obtained by research funds allocated by heads of departments. A spectrum of mechanisms for distributing funds should be pyramidal, with a base of many project grants and relatively few large interdisciplinary research centres. Though the government is now considering separating the allocation of funds for teaching and research in Britain, the committee warns against taking any responsibility for research funds away from the universities.

Christine McGourty

NEWS IN BRIEF

Alaskan oil spill

Washington

THE oil tanker *Exxon Valdez* last week struck a reef and spilled some 11 million gallons of crude oil into Prince William sound off the south coast of Alaska. The spill is being called the worst in North American history. The tanker had just left the port of Valdez, the south terminus of the Trans-Alaska Pipeline and was carrying 53 million gallons of oil. Clean-up crews are trying to contain the spill which threatens severe disruption to the local ecosystem. J.P.

Better health ahead

Sydney

A NATIONAL strategy for Aboriginal health, released at the annual meeting of health ministers last week, is the first attempt in ten years to deal with the health problems of that population. Australia's 228,000 Aboriginals suffer from poorer health and a lower life-expectancy than the general population. The World Health Organisation said last year that Aborigines have the worst reported standard of health of any indigenous people. The government's strategy includes plans for upgrading health services, providing more health training and educational programmes, and transferring basic Aboriginal health services from the state governments to the Aboriginal health community. The recommendations, if implemented, are expected to cost A\$417 million. T.E.

Hair-raising scheme

New Delhi

HUMAN hair is one of India's resources that has heretofore gone to waste. But an Indian and a Japanese company have now formed a partnership to turn what was discarded into profit.

Hair is a potentially profitable source of amino acids. Under the new arrangement, Union Bros of Japan will supply expertise, equipment and part of the finance for a \$10-million processing plant being set up at Pondicherry in south India. The plant will consume 1,200 tonnes of human hair a year, turning it into 80 tonnes of L-cysteine, 24 tonnes of L-tyrosine and 1,400 tonnes of dry powder containing as many as 19 amino acids. The products are extensively used in the pharmaceutical, food-processing and cosmetic industries. The company expects to make 25 per cent profit and Japan will purchase the entire output for 8 years. Human hair has become an important raw material for amino acids after whaling restrictions made that source scarce. Apart from barbers' shops, the biggest source of human hair in India is the temples, where Hindus shave their heads as a religious custom. One temple in Tirupati Inandhra Pradesh collects 4,000 tonnes of hair a year. K.S.J.

Cold results from Utah

Washington

AN announcement last week by the University of Utah in Salt Lake City that physicists there had achieved nuclear fusion in a test tube at room temperature has been followed by a flurry of speculation about not only the strange science involved but also the peculiar publicity given to it.

The technique involves nothing more complicated than electrochemistry. Current is passed through a palladium electrode immersed in a solution of various salts in deuterated water. Over a period of many hours, deuterium atoms are absorbed by the palladium lattice in such numbers that some of them fuse, forming tritium or helium-3, simply because of their proximity. A return of four watts of heat from one watt of electricity was reportedly obtained when current was passed for a total of 100 hours.

According to James Brophy, vice president for research at the University of Utah, the decision to announce the work in preliminary form at a press conference last Thursday, before a paper had been accepted by a scientific journal, was a response to press enquiries. But the university press office had previously contacted several science reporters and encouraged them to attend.

A statement by the University of Utah that results "will appear in the scientific literature in May", coupled with a report in last Thursday's *Wall Street Journal* that papers had been submitted to *Nature*, provoked numerous telephone calls to *Nature*'s Washington office enquiring whether we had accepted or were about to publish one or more papers. (The answer is no.) Meanwhile, a group at Brigham Young University in Provo, Utah, led by Steven E. Jones, has refused to say whether they have obtained similar results.

Enthusiasm over the claim has been tempered with a good deal of scepticism, not only due to the lack so far of a detailed account but also because the history of fusion research has been regularly punctuated by 'breakthroughs'. The new claim bears some similarity to muon-catalysed fusion, in which the substitution of a muon for an electron in heavy hydrogen shrinks the size of the molecule enough for spontaneous fusion of the two deuterons to occur. Reasonable rates have been achieved in small-scale experiments, but as an energy-producing method muon fusion is no nearer success than any other technique. The University of Utah has applied for a patent on its technique, but Jones says only "don't sell your oil stocks yet". □

Academy starts over again

Moscow

AT the general meeting last week of the Academy of Sciences of the USSR, only 8 of 20 delegates were elected to the All-Union Congress, from which delegates will ultimately be elected to the Supreme Soviet of the USSR. The Academy's election process will now have to begin again.

The normal election procedure was complicated by the fact that after an extended plenum of the leading bodies of the Academy on 18 January, just 23 candidates for the 20 delegates out of 130 candidates proposed by research teams at academic institutions were chosen (see *Nature* 337, 293; 1989). The failure to nominate those scientists receiving the greatest popular support—including Academicians Andrei Sakharov and Roald Sagdeev—evoked particular discontent. A lobby was set up for the democratic election of People's Deputies from the Academy of Sciences of the USSR, which held a protest meeting in front of the Academy Praesidium's building in Moscow, and organized what amounted to a boycott of the election of the nominated candidates.

According to the law, additional elections have to be held to fill the remaining 12 delegates within the next two months. The new election process began during last week's general meeting, when procedures and potential candidates were discussed. Describing the outcome of the elections in his speech at the conference, Sakharov emphasized that emotions were running so high not because "Sakharov was not nominated, but because the role of the research institutions was ignored".

Academician G.I. Marchuk, president of the Academy, explained the current election phenomenon by saying that "all of Soviet society, including the Academy, is going through a historical moment—the transition to people's power has started in the USSR in earnest. The main point is that the elected deputies will represent not just the interests of the Academy but also and above all the interests of the social development of the nation as a whole."

Voting procedures now call for the 15 unsuccessful candidates to be asked to agree to run again. In addition, many scientists from the previous list suggested by various institutes, as well as new names, were mentioned. Judging by the response of the audience, the following scientists should be on the list of new candidates: Academician Stanislav Shatalin, an economist; Academician Vitaly Ginsburg, a physicist; and the economist Nikolai Shimeley. New elections will probably be held next month.

The conference decided that new candidates will be nominated by the Praesidium of the Academy of Sciences of the USSR. This responsibility has now been vested in it after the extended plenum of the leading bodies of the Academy failed to rise to the occasion. The members of the former election commission of the Academy of Sciences on the election of People's Deputies have resigned, and a new commission will be formed. The date for a new election conference has not yet been fixed. Importantly, the voting will be by secret ballot at meetings of research associates of academic institutes.

Yuri Kanin
Novosti

WEATHER FORECASTING

Precipitation precipitates complaints

Sydney

AN insufficient number of weather stations are being blamed for Australian weather forecasters' failure to provide early warning of torrential rains that flooded parts of central Australia, causing significant stock losses. Rains have been particularly heavy in Australia this year. At Broken Hill, in western New South Wales, 180 millimetres of rain fell in a slightly more than 24 hours, the heaviest fall that had been recorded since 1973. Storm warnings by the Department of Meteorology came on the morning of the fall, too late for some farmers.

A reduction in staff salaries and productivity costs means much of Australia weather remains undetected. "We have a country the size of the United States with one tenth its population. The quality of

our meteorological services is related to the observational networks we can afford", said Bill Downey, assistant executive director for the Australian Bureau of Meteorology.

Peter Noar, assistant director of weather services, acknowledges that a full network of a couple of hundred observation stations is needed. Some help may come from plans to expand land-based automatic weather stations and improve the network for aviation forecasting.

But the problems at Broken Hill are threatening to get worse before they get better. The Civil Aviation Authority airport observation station may close, forcing the Bureau of Meteorology to try to come up with sufficient funds to open its own station in Broken Hill.

Tania Ewing

Banging the Freedom drum

Washington

NATIONAL Aeronautics and Space Administration (NASA) officials like to say that no matter what objectives are chosen by the United States for space activities in the twenty-first century, "all roads begin with the space station Freedom". But as NASA pleads its case before a budget-conscious Congress, Freedom's \$16,800-million price tag practically guarantees that its own road into space will be a rocky one.

Now set to be deployed in the mid-1990s, the space-station programme is building momentum. Last September, Canada and the European Space Agency signed formal participation agreements (see *Nature* 335, 480; 1988), and earlier this month Japan signed a memorandum of understanding. Altogether, NASA's foreign partners will contribute \$8,000 million to the project.

President George Bush is an enthusiastic proponent of the space station. Nervous about the cost, Congress last year appropriated \$900 million for the space station, but ordered \$515 million to be held back until 15 May this year so the new president could decide whether to proceed. President Bush has already indicated that the money will be released for the space station.

This year, NASA is requesting \$2,050 million for its space-station activities. Three-quarters of that money will go to the four aerospace companies with which NASA signed contracts last September to build major elements of the station. Boeing is developing the crew living quarters and logistic elements; McDonnell Douglas is pursuing the central truss structure and resource nodes; General Electric is developing the polar platform and attached payload accommodations on the main station; and Rocketdyne is developing the power systems.

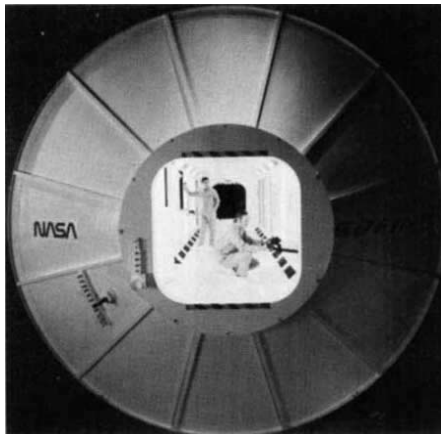
For its part, Congress is enthusiastic about the idea of the space station, but budgetary concerns make any new and expensive project vulnerable. The word around Congress is that as much as \$1,000 million may be cut from NASA's total 1990 budget request of \$13,300 million, and some of that is bound to come from the space station.

But space-station officials are already feeling pinched, and have warned that further cuts could be devastating. NASA has not been able to find funds for nearly a third of the 2,644 posts authorized for the programme. At a hearing on 16 March before the Senate Commerce, Science and Transportation Subcommittee on Science, Technology and Space, NASA Associate Administrator for the Office of Space Station James Odom

testified that "virtually any cut would do major hurt, and a significant cut would be tantamount to killing the programme". Odom also warned that backing out on NASA's commitments to its foreign partners would also have devastating effects.

Launching the space station is planned to take 20 space-shuttle flights, the first of which is now scheduled for 1995. Human occupancy should be possible after the fourth shuttle flight, and a permanent crew can stay on board after flight 13.

The space station's attached science payloads — now likely to include a gamma-ray/cosmic-ray focusing device called Astromag and a cosmic dust sampler — are intended for flight 7, an early



Installing equipment in Boeing's mock-up of Freedom's living quarters.

flight that supporters say is indicative of NASA's commitment to doing science on board the space station. The launch schedule for the flights is based on current shuttle lift capabilities, and a new advanced solid rocket motor, once it becomes available, could speed up the process.

A decision to move forward with shuttle-C, a cargo-only version of the shuttle devised as an alternative option, would also be a boon for the space station, but so far the administration has shown little support for this plan.

Coordinating the development of hardware with that of scientific instruments so that both will be ready to fly on time will be a tricky business. As an example, General Electric has already begun work on the polar platform, but funding for the platform instruments, which have already been selected, will not be formally requested until the 1991 fiscal year. Last-minute modifications to spacecraft hardware to accommodate instruments can be extremely expensive, and can cause delays.

Although NASA has agreed with the

On-again, off-again Landsats on again

Washington

LANDSATS 4 and 5 are back in business — for the moment. The Earth Observation Satellite Company (EOSAT) announced on 15 March that the shut down order issued by the National Oceanic and Atmospheric Administration (NOAA) had been rescinded.

EOSAT, which operates the two Earth remote sensing satellites for NOAA, had begun shut down procedures that would have terminated Landsat operations at the end of the month (see *Nature* 338, 194; 16 March 1989).

Details of the financial rescue package have not been revealed, but EOSAT says it costs \$9.4 million to operate the two Landsat spacecraft. NOAA officials will only say that the reprieve will last at least until the Bush Administration completes a review of the programme now under way.

The rescue effort was coordinated by the new National Space Council chaired by Vice-President Dan Quayle. Members of Congress have reported that their offices have been inundated by letters asking that money be spent to keep the Landsats alive.

Joseph Palca

White House Office of Management and Budget to cap spending on the space station at \$16,800 million, several elements of the station, such as operations costs, facilities, launch costs, communications support and several new features including an emergency crew escape vehicle, are not covered by the cap. Other programmes at NASA have been cut back in order to provide adequate money for necessary space-station development and activities.

There has long been tension between space scientists and those favouring manned space activities. Station chief scientist John Bartoe acknowledges that selling the station solely as a tool for science is untenable, but that it can be useful for scientists once they learn its capabilities. It seems likely that the successor to James Fletcher, who has announced his resignation as NASA administrator, will continue the strong support NASA has traditionally provided to the manned space programme.

Proponents of the space station are trying to secure the high ground in their efforts on behalf of the space station. At the Senate hearing earlier this month, Senator Al Gore (Democrat, Tennessee) set the tone for the coming Congressional debate that will make opponents of the space station squirm: "The choice facing us this year is not just between moving ahead with the space station or not, but between continuing leadership in space and abdication and decline."

Joseph Palca

Transgenic sticky issues

Washington

ANIMAL rights, the release of genetically engineered organisms into the environment, and the effects of large agribusiness companies on the farming industry are the real issues behind the debate over the patenting of animals, according to a report* released last week by the US Congress's Office of Technology Assessment (OTA). Existing regulations can be adapted to address most of the practical considerations of animal patenting, such as whether farmers should pay royalty fees for breeding patented livestock. But the ethical question of whether or not transgenic animals should be subject to patents is a question that may need further clarification, OTA says.

Congress has heard both sides of the debate since the patent office declared in early 1987 that it would not reject patent applications on the sole basis that the invention for which the patent was being sought was an animal. The granting of the first patent for a higher life form last year — to Harvard University, for a transgenic mouse containing human on-

cogenes (see *Nature* 332, 668; 1988) — heated up the exchange between patent proponents and those who asserted that animal patenting would lead to animal suffering and higher costs for farmers.

But attempts in the past two years to pass legislation to halt the patenting of additional animals until the ramifications could be worked out — and specifically to exempt researchers and farmers from royalty payment requirements — have floundered. The Animal Legal Defense Fund has also lost the first round in a court battle to overturn the patent office's decision to patent animals.

In the meantime, 44 patent applications covering animals have piled up at the patent office, and the exclusive



Du Pont's OncoMouse, the only patented animal so far.

licensee of Harvard's patent, Du Pont, is taking orders for its \$50-apiece OncoMouse, which it plans to begin shipping in May. Other companies are not waiting for patent protection to develop profit-making genetically engineered animals. Integrated Genetics has formed a collaboration with Tufts University to develop herds of cows capable of excreting human proteins in their milk, and the company Transgenic Sciences has just been formed, and intends to develop animal breeds which produce valuable pharmaceuticals.

Representative Robert Kastenmeier (Democrat, Wisconsin) last week reintroduced bills he put before Congress last year that would create a transgenic animal advisory committee within the Department of Agriculture which would authorize permits to field-test transgenic animals. Kastenmeier's bills also specifically exempt farmers and researchers from paying royalties on patented animals. Kastenmeier says he hopes biotechnology companies and opponents of animal patenting can work out their differences, an acceptable compromise, but

New carp park ahead

Washington

THE biotechnology review board of the United States Department of Agriculture last week approved the first field-test involving a genetically-engineered fish. Researchers at Auburn University in Alabama have been given permission to introduce carp which contain trout growth hormone genes into a test pond at the university.

The carp were developed through a collaboration between researchers at Auburn, and at Johns Hopkins University in Maryland. Carp containing the trout genes have been shown in the laboratory to grow significantly larger than normal carp, a finding that could have commercial significance for fish farming.

The carp will be kept in a contained pond, with a series of barriers to prevent their escape into open water. As an added precaution, a mechanism has been developed to introduce poison into the water should a fish manage to get beyond one of the barriers.

The Auburn carp is the first transgenic animal to be approved for release into the environment by the Agricultural Biotechnology Research Advisory Committee, which has been charged with developing guidelines for reviewing field-test applications from researchers working under Department of Agriculture grants. The finishing touches are now being put to the committee's guidelines, which should be published in the Federal Register for public comment within the next two months.

The Department of Agriculture is opposed to legislation introduced last week by Representative Robert Kastenmeier (Democrat, Wisconsin) which would require it to set up a separate review board to grant formal permits for field-tests of transgenic animals. Daniel Jones from the Department's Office of Agricultural Biotechnology says that laws already on the books should cover transgenics, with some "fine-tuning" from the new guidelines. Jones says the guidelines will parallel those covering recombinant molecules developed by the National Institutes of Health: they will be compulsory for academic grantees, and industrial researchers are expected to adhere to them voluntarily.

Carol Ezzell

that he will support a halt to animal patenting if the means to address the sticky issues surrounding such patents is not put in place.

The Congressmen who proposed a moratorium on the patenting of animals last year, Representative Charlie Rose (Democrat, North Carolina) and Senator Mark Hatfield (Republican, Oregon), have no plans to introduce their bills again this year.

Carol Ezzell

CHEMISTRY

NERC finds British chemistry ailing

London

CHEMISTRY in Britain is suffering from insufficient investment, inadequate equipment and poor dissemination within the research community of information on facilities available. Those are the main conclusions of a committee set up by the Natural Environment Research Council (NERC) to review the health of chemistry. The report, published this week, concludes that facilities are outdated, and most researchers feel they are lagging far behind their counterparts in other European countries and in the United States. The committee says that because much of the equipment in universities and NERC institutes is old, slow and insensitive, the quality and quantity of data is suffering. This in turn may be leading to the loss of commissioned research contracts.

To remain competitive, Britain must have well-equipped, state-of-the-art laboratories, and this will require a substantial programme of investment in chemical equipment, says the committee. Its report points out that although the cost of equipment is increasing faster than the rate of inflation, its speed and sensitivity are increasing even more rapidly, so the cost-effectiveness of new equipment is improving in real terms. At present, even the equipment available is used insufficiently because of severe constraints on travel and subsistence budgets.

Christine McGourty

*New Developments in Biotechnology — Patenting Life, Office of Technology Assessment, Washington DC, 1989.

CHERNOBYL

Soviet data made public

London

RADIOACTIVE contamination from Chernobyl is still "an acute technical and social problem" likely to persist for a considerable time, according to Dr Yuri Izrael, chairman of the State Committee for Hydrometeorology, writing in *Pravda* last week. In a full-page article, he presented the first figures and maps for the whole area of the Soviet Union affected, comprehensively extending the data for Byelorussia given at a public meeting in Minsk last month (see *Nature* 337, 683; 28 February 1989).

The release of these data is an important development, a departure from the policy described by Kiev radio earlier this month as the "tranquillizing" of the population. Contrary to the impression conveyed at Minsk that the collection of data had only just been completed, Izrael also says that the collection of data began on 26 April 1986, the day after the accident, and that a generalized map covering a 60- by 40-km region showing contours of gamma-radiation was available on 10 May that year, in time to guide the evacuation of people from contaminated zones.

In the first few days after the accident, Izrael says, the first 2 milliroentgen per hour contour enclosed about 200,000 km² of Soviet territory, with islands of contamination as far away as the Kola Peninsula in the far north, Ivano-Frankovsk in the south-west and the Caucasus in the south-east.

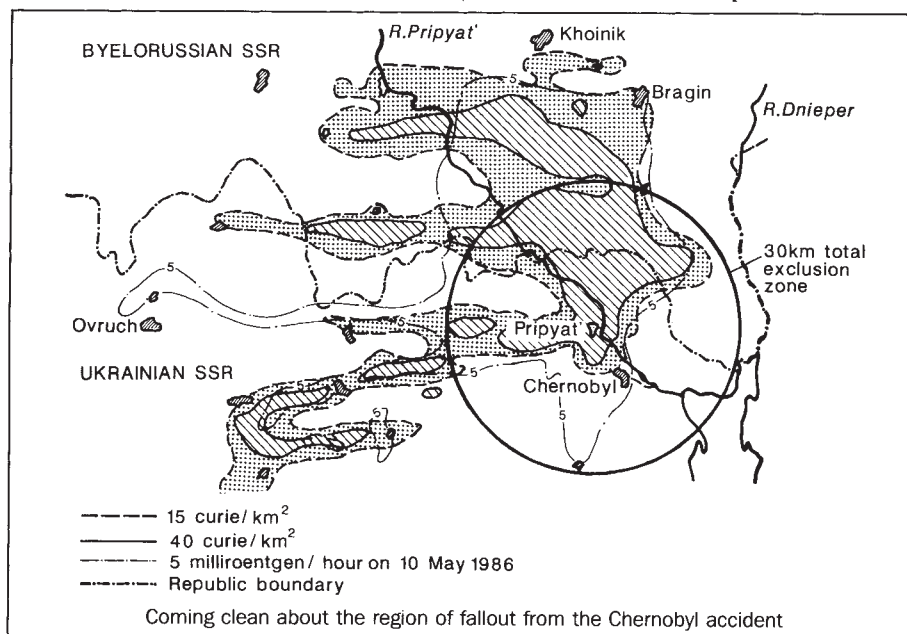
"Massive" testing of soil samples began in June 1986, when exceptionally high proportions of the long-lived isotopes caesium-134 and caesium-137 were found. But Izrael says that the lifetime exposure limit of 35 rem of external radiation was applied to the caesium-

contaminated areas only in November 1988, leading to the evacuation of a further 20 villages in Byelorussia.

Pollution of rivers and reservoirs remained "within the limits of the agreed norms" in the months after the accident, according to Izrael, who says that total beta-activity amounted to between 1 and 6×10^{-9} curie per litre during the first two months. (The safety limit was 10^{-8} curie per litre.) During the spring floods of 1987, Izrael says, even the rivers with radiation 'hotspots' in their catchment areas were well below the limit.

Further aerial and ground surveys carried out in 1988 are said to have revealed substantial changes of the contours. At present, Izrael says, for caesium-137 the 15 curie per km² contour encloses some 10,000 km², of which some 3,500 km² have been evacuated, while the remainder has a population of 230,000 people. In the town of Slatytych, built for Chernobyl workers after the accident, radiation levels (0.013 to 0.03 milliroentgen per hour) are close to the natural background, although there are some hotspots in the surrounding forests which nevertheless do not exceed the safety limits.

Although Izrael's sympathy for the new anti-nuclear groups in the Soviet Union is clearly muted, he concedes that the decision to close the two Armenian reactors (at Yerevan, one last month and one this) was "quite understandable". But he concludes that, with the threat of the greenhouse effect, nuclear generation is the "most promising" form of energy. The lessons of Chernobyl, he says, imply that the operation of existing nuclear reactors is "many times", and of future stations, "many orders of magnitude", safer than in the past. **Vera Rich**



ABORTION

French drug under attack

Paris

THE controversial abortion pill RU-486, which was approved by the French Health Ministry for use in abortion clinics last year amid angry opposition, is again under attack (see *Nature* 336, 4; 1988). Dr John Willke, of the US National Right to Life Coalition, is in Europe for the annual meeting of the International Federation for the Right to Life in Brussels. Before the meeting he met the medical research director of Roussel-UCLAF, the company that makes the drug, and suggested that the company's products could be boycotted in the United States if the drug is not withdrawn.

France is the only country to have legalized the use of RU-486 for abortion, and the drug is administered only under strict medical supervision and only in association with prostaglandins. In October last year, Roussel-UCLAF suspended sales of the drug after some of its senior employees and their families received anonymous death threats. It was only after government intervention that the drug was again put on the market.

Roussel-UCLAF, a subsidiary of the pharmaceuticals giant Hoechst, declined to comment on the meeting with Willke. A spokesman said that RU-486 is only a new alternative to surgery and its withdrawal would have no effect on the number of abortions carried out. In Washington, the National Right to Life Coalition also declined to comment on possible boycott action while Willke is still in Europe.

Right to Life is one of the most vocal and powerful minority lobbies in the United States, capable of assembling thousands of people at rallies to protest against abortion. Any decision to boycott Hoechst products would therefore be taken seriously by the company, even though it has no plans to introduce RU-486 in the United States.

Although it is illegal for federal laboratories to carry out research on abortifacients, the potential contraceptive properties of RU-486 are being studied. At low doses, the drug is thought to impair the implantation of a fertilized ovum in the uterus lining. Since implantation is a criterion for pregnancy, such a function would technically be 'contraceptive' and not 'abortifacient'.

There is no national legislation on abortion in the United States; only a Supreme Court ruling makes it illegal for states to outlaw abortion. But a test case currently before the Supreme Court has challenged this ruling. The judgement, which could come in the next month, will stir emotions in the anti-abortion lobby, whatever the outcome. **Peter Coles**

Hope for nuclear industry?

Boston

THE US Nuclear Regulatory Commission (NRC) is expected soon to authorize sweeping regulatory changes that will shorten the licensing procedure for nuclear plants. The new procedure eliminates the requirement for a separate operating licence for newly constructed reactors, thereby limiting the opportunity for public intervention once a plant has been built. The five-member nuclear commission is expected to vote on the licensing overhaul as early as this week, and the measure is said to be almost sure to win approval.

Renewed interest in nuclear power and frustration over the experiences at the Shoreham and Seabrook plants —

both now built but not operating — have propelled consideration of the new rules. Proponents of the changes argue that the current licensing procedures are unnecessarily cumbersome and expensive to follow.

If adopted, supporters say the new rules will shorten the time needed to begin operating a new plant, simplify the site-selection process and bring standardization in reactor design. In addition, advocates believe that the changes would be an important symbolic incentive to the nuclear industry and to poten-

ess. The new rules would not affect any existing plants (licensed or otherwise).

The idea of revising the licensing procedure for nuclear power plants has been considered for several years, but has usually been discussed as a matter to be dealt with by Congress. Although members of the nuclear regulatory commission and the nuclear industry have urged Congressional action to simplify the licensing procedure, Congress has so far failed to take up the matter. According to one inside observer, the commission "finally got fed up with Congress and decided to go out and change the licensing procedure on their own".

But Ken Bossong, director of the Crit-

USCEA

INTERNATIONAL SCIENCE

Non-aligned line-up for new technology

New Delhi

A centre for science and technology of non-aligned and other developing countries was formally established in New Delhi last week after 31 countries signed the centre's statute. Joint research programmes and exchange of experts are some of its objectives, but the centre primarily seeks to pool existing scientific resources, technological and consultancy capabilities in member countries for their common benefit. The centre will be financed by contributions from each member country, fixed at \$15,000 per year. The initial 31 countries are from the Indian subcontinent, Africa and Eastern Europe, but China is not a member. As the host, India has offered an extra \$1.1 million for office building, equipment and transportation. The first-year budget of the centre is \$1.6 million. It will have an Indian director but other key posts are to be filled by people from other member countries.

Computer software, biotechnology, renewable sources of energy telecommunications and new materials are among the areas from which about 30 specific collaborative programmes have been identified. Project expenses will be met by participating countries; the centre will play only a catalytic and coordinating role, although there will be money for pilot plants or infrastructure development.

The centre's first task is to prepare a directory of science and technology institutions, and a list of transferable technologies already available in member states. There are proposals for a common gene bank, a shared database on science and technology and a consortium for technology promotion within the non-aligned club.

K.S. Jayaraman

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The way things were: work in progress on the Palo Verde nuclear generating station

tial financial backers of nuclear power.

But critics argue that the proposed changes are illegal and unsound, and that they are a "blatant effort" to limit public involvement in the licensing process. If passed, the rule changes are sure to face a legal challenge. According to several sources, anti-nuclear groups are meeting currently to decide "who will take the lead" in the legal challenge.

Under the proposed licensing process, only one licence will be given by the nuclear regulatory commission at the construction phase of a new nuclear reactor which would be a combined "construction and provisional operation" licence. At some point shortly before construction of a plant was complete, the utility would seek approval to 'implement' its operating licence.

In other changes would allow utilities to obtain "early site permits" up to 20 years before building a plant. Public disputes over the suitability of the location for the reactor would have to be settled at this early stage. The nuclear regulatory commission would also "pre-approve" several standardized nuclear reactor designs. By using one of these designs, utility would be exempt from part of the present public review proc-

ical Mass Energy Project, likens the proposed changes to awarding a university students a degrees "based only on their first year of school". He and other critics contend that the proposed changes "decrease if not eliminate" the possibility to consider flaws or fraud in the construction process, or changes in the site situation — such as roadways, population changes, hospitals or schools — that may occur during the 20 year period after the site has been approved.

In addition, Bossong and others say that the new rules preclude evaluating the financial ability of the utility to operate the plant once construction is finished. Despite the controversy already raging around the rule changes, there is little certainty over how they might affect the beleaguered nuclear industry.

Although they ought to make the construction and licensing process simpler and more concise, several observers note that the financial community is greeting the proposed changes warily. As Bossong puts it, "the industry still has to grapple with widespread public opposition to nuclear power, lack of nuclear waste facilities and economics that are not competitive with other sources of energy".

Seth Shulman

GREEN POLITICS

New pragmatism pays off

Munich

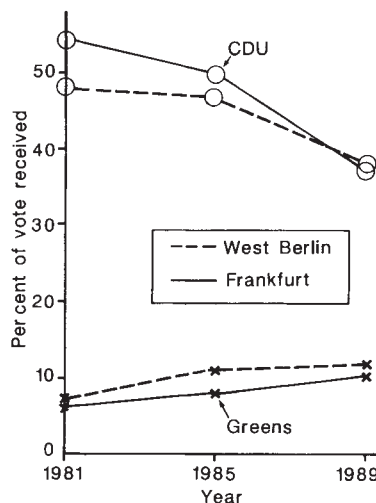
THE Greens are on the rise again in West Germany. For the first time in two years, a *Land* (state) government including the Green party was formed in West Berlin on 16 March, pointing the way towards more Green political power. Ten days earlier, national party members elected an executive committee (*Vorstand*) intent on forming a coalition with the Social Democrats (SPD) in the *Bundestag* after the 1990 election. The possibility of an SPD/Green coalition is stronger than ever in the aftermath of heavy losses for the conservative Christian Democratic Union (CDU) in recent elections.

Even as their political fortunes are waxing, the Greens are taking a less utopian view toward ridding society of what they see as its technological evils. A survey of Green politicians in Bonn and elsewhere reveals a more pragmatic attitude towards divisive issues such as genetic engineering, nuclear research and space exploration. Once advocates of banning everything to do with technology, ten years of political experience has tempered Green extremism.

The Greens shared power once before in a stormy coalition with the SPD in the *Land* of Hesse from 1985 to 1987. But a conservative coalition led by the CDU took over, in part because of governmental paralysis on nuclear issues. They faded from view after that, plagued by internal struggles. The election of a pragmatic leadership at the Green party conference in early March reveals the new direction of the party. The three executive committee spokespersons elected—Verena Krieger of Witten, Ruth Hammerbacher of Osnabrück and Ralf Fücks of Bremen—were careful to avoid the kind of name-calling that was so destructive two years ago. The election, and the

decision of West Berlin's Greens to form a coalition with the SPD, reflects the wishes of Green voters to see their leaders take action, even if it means compromise.

Because of its unique status, the Greens' success in West Berlin is hard to extrapolate to the rest of the country.



Small but steady progress of Greens.

The *Alternative Liste* (AL), a local Green party, had always been strong in some areas of West Berlin, which has a large population of draft-resisters and dropouts from society. After weeks of debate, the SPD and AL reached a compromise on governing the city.

But the results of community elections in Hesse on 12 March suggest that the Berlin result was not a fluke. Green support surged all over Hesse and helped topple a CDU government in West Germany's financial capital Frankfurt. A new political constellation seems to be forming which is widely seen as a threat to the re-election prospects of Helmut

Kohl's conservative coalition in 1990. According to a poll taken by the magazine *Spiegel* in February, Kohl's CDU and its Bavarian sister party CSU had reached a two-year low in popularity.

The compromise worked out between SPD and AL in West Berlin gives some clue as to what sort of science policy to expect in a national SPD/Green coalition. For example, the AL yielded on initial demands to stop construction of a nearly completed nuclear research reactor at the Hahn-Meitner Institute. Receiving a final licence for the reactor from the new environment senator, a member of the AL party, will be difficult, but the SPD is expected to ensure that it will be possible.

National Green politicians are quick to assure that they would not shut down research institutions. "It would be crazy to forbid nuclear physics research", says *Bundestag* member Otto Schily, "just because the atomic bomb is a danger to mankind". If the Greens come to power, says Schily, the freedom of research that is guaranteed in the West German constitution (*Grundgesetz*) "must not be touched". Schily belongs to the more 'realistic' wing of the Greens.

The most visible feature of the West German research scene that would be affected by Green involvement in government is space policy. West Germany's active collaboration in European Space Agency projects has been supported by a publicity-conscious government that seeks as much prestige as possible for its space programme.

Fücks criticized the space programme because of its ties to the military. Given all the ecological goals the Greens have in mind, it would be hard to continue to invest in expensive programmes such as manned space flight. Despite the criticism, it is unthinkable that a new West German government would renege on international contracts signed by its predecessors. Schily categorically excludes this possibility. The Greens still advocate a shutdown of all nuclear plants, and their views are not far from those of the general public in West Germany. Nearly 80 per cent of the population would prefer an immediate or gradual shutdown, according to the *Spiegel* survey.

Like nuclear power, genetic engineering is an issue that calls for sweeping condemnations from Green politicians. But the current leaders have begun to break from their predecessors in admitting some practical uses for molecular biology, such as the development of AIDS treatments.

Maintaining their political momentum going into the long series of *Land* elections in 1990, culminated by the *Bundestag* election, will be difficult. But recent events have ensured that the Greens will not just go away. **Steven Dickman**

ENVIRONMENTAL PROTECTION

Compromise reached on toxic waste

Paris

DELEGATES from 116 nations meeting last week in Basel, Switzerland, under the aegis of the United Nations Environment Programme, failed to reach unanimous agreement on rules governing dumping of toxic waste. Only 34 nations signed a treaty, which had taken 18 months to draft, but 105 did sign a final compromise document. The United States, Britain and the Soviet Union were among those calling for more discussion before signing a treaty.

Some developing nations wanted a total ban on the export of toxic waste because, they argue, they are easy prey for industrial companies seeking to dispose of waste away from their own back yard. But for most European and other industrialized nations, waste disposal is an acceptable

commercial enterprise and a ban would be out of the question.

The Organisation for African Unity sought powers to search ships in coastal waters for clandestine cargoes of hazardous substances, but this is against international maritime law. As a compromise, the summit drafted an agreement by which signatories undertake to export waste only to countries having sufficient facilities and expertise to deal with it and only following written confirmation by both the exporter and the importer. Signatories also agree to reimport waste rejected by another country, but they will not be obliged to accept responsibility for accidents arising from waste dumped before the new regulations come into force.

Peter Coles

Distinguishing fraud from error

SIR—There is a danger in the controversy over fraud in science of merging the concepts of fraud and error. The call for an audit of scientific papers for error is a symptom of this trend. Fraud such as fabricating data or publishing the work of others as one's own is of course serious, particularly when it involves assessment of drugs and other medical treatments where lives are at stake. But error is an inevitable part of science. This fundamental point is being missed in the current debate. Scientists struggle to express their ideas, struggle with recalcitrant organisms, use advanced laboratory techniques that they barely understand, and study systems so complex that many hypotheses may seem to fit the data. This is not the orderly process that some imagine science to be. One does not go into a laboratory and order a cure for heart disease and get a guaranteed result. Results in science are never guaranteed and are never free from error. Newton, Einstein and Darwin were all wrong in some respects, even though they are the standard by which we judge genius itself.

Science is a complex and subjective process, one man's enthusiastic promotion of an hypothesis may be another man's fraud. The history of science is full of individuals who turned out to be right but who, by the strict criteria proposed by some today, could be attacked for fraud. In the laboratory notebooks of Millikan's oil-drop experiments, it can be seen that he rejected many trials because his judgement told him that something was wrong. In today's climate he could be accused of blatant manipulation of the data, that is, fraud. The 'discoverer' of N-rays had bad technique and practised self-deception, but there was no fraud. Was Columbus guilty of scientific misconduct for his self-delusion about the circumference of the Earth? Kepler defended himself in his search for the "Music of the Spheres".

Some of the charges levelled in recently publicized cases include misclassification of control subjects, poor laboratory procedure (Benveniste) and complex issues of scientific method (Baltimore). These are real scientific issues, but not causes for hand-wringing and raised eyebrows, much less for congressional scrutiny.

Who ever said that when you open a journal you are reading a message carved in stone by some higher being? Every time one reads the scientific literature, it is necessary to be wary and to compare, test and evaluate methods, data and statistical methods. At no point do we have such a grasp on the truth, statistical, experimental or otherwise, that a panel can be certified as error hunters. Reviewers of a

paper are themselves often wrong (the original theory of continental drift was met by howls of laughter from 'respected' scientists). Who will review the error hunters? Who is qualified to punish whom?

So, is there error? Of course there is. But we should also note that every building built has flaws and half of them are ugly, every economic policy ever formulated is imperfect. We can only ask for perfection when the correct answer is known (for example, on the factory assembly line), but if the answer is known then it isn't science.

CRAIG LOEHLE

*Environmental Sciences
Division, 773-42A,
Savannah River Plant,
Aiken, South Carolina 29808-0001, USA*

Defence research

SIR—You report (*Nature* 336, 610; 1988) that the veterinary school at this university is being supported in a research project on the survival of bacterial pathogens in aerosols by the Ministry of Defence (MoD) Chemical Defence Establishment. The project is of evident relevance to biological warfare, if for no other reason than that proposals must have 'defence relevance' to qualify for MoD support. Whether or not the ministry's interests in biological warfare are legitimate, we believe that they are not appropriate to a veterinary school.

Unlike the research councils, MoD has no remit to improve animal health and welfare nor a broad peer-group-defined strategy. The medical and veterinary benefits of the Bristol project are merely a spin-off of its defence interests.

While research support by the research councils appears to be increasingly inadequate, MoD support of research in British universities and higher education establishments has increased by 60 per cent since 1984-85 (from £10 million to £16 million). Consequently there is a danger that scientists are relying increasingly, as at Bristol, on support from government or from commercial organizations with narrow research objectives and a strong interest in confidentiality, even secrecy. We believe that such support undermines the integrity and openness with which research can be conducted.

Clearly, the freedom with which a project is formulated is compromised when it has to conform to concerns of a sponsor (for example, defence or financial profit) that are completely irrelevant to the purpose of the research (for example, the improved health of livestock). The freedom to publish is not adequate safeguard of independence when the freedom to propose is restricted.

We urge research institutions and individuals to take a long hard look at the way in which the motives of funding bodies

may subtly redirect and compromise research priorities and independence.

PETER DUKES, SUSAN MAYER, ALISON BLAXTER, HELEN BROADHEAD, MARK J. NEWTON-CLARKE, BEVIS MILLER, KAREN STEVENS, A. DOUGLAS WILSON, PAUL R. WOTTON, JOSEPH J. McNAMARA, ESBJORN TELEMO, MARTIN WARNES
*School of Veterinary Science,
University of Bristol,
Langford House, Langford,
Bristol BS18 7DU, UK*

Imperial Japan

SIR—We should like to comment on the article (*Nature* 337, 107; 1989) in which Japan's deceased Emperor Hirohito was described as an admirable biologist. We fear that the article may be abused by those who wish to enhance or maintain the emperor system, and that it is likely to be cited in sentences such as 'A leading scientific journal, *Nature*, also admires the Emperor as a prominent biologist'.

Since Hirohito died on 7 January, programmes on television and articles in newspapers and magazines have been full of admiration and glorification for the personal qualities of the emperor. Neither the mass media nor leading intellectuals have touched on the dangerous and inadequate aspects of the emperor system, which have been used to control the ideology and attitude of Japanese citizens. Opinions against the emperor or the emperor system are suppressed by both direct and indirect pressure.

There are some problems in associating biology with the emperor. First, the words 'naturalist' or 'biologist' give a relatively good impression, and have been used to make an admirable personality of the emperor. Second, there are few faculty positions and departments in the fields of fundamental biology (except biotechnology and molecular biology) and funds are also scarce in Japan. But biologists can get money for international meetings, and other scientific activities in subjects related to those in which the emperor was interested. So some biologists have tended to use the emperor to get funds, while the government and conservative intellectuals have used such biologists and their activities to enhance the respectable image of the emperor. The International Biological Prize (established to celebrate the sixtieth year of the emperor's reign) is also used to enhance the image of Japan's emperor in the eyes of other countries.

MASAKADO KAWATA

*Department of Biology,
Faculty of Education,
Shizuoka University,
Oya, Shizuoka 422, Japan.*

EIICH KASUYA

*Laboratory of Biology,
Faculty of Education,
Niigata University,
Ikarasi, Niigata 950-21, Japan*

Rise and fall of Western diseases

David J. P. Barker

Nutrition and hygiene during early childhood are important in determining the risk of the diseases which follow industrialization. We still know little about the processes involved.

THE term 'Western diseases' is used to describe a group of diseases common in industrialized countries but uncommon elsewhere¹. These diseases are regarded as a consequence of the environmental changes accompanying industrialization and prosperity; it is often argued that their prevention depends on a return to practices of the past — for example, resumption of a diet high in complex carbohydrates and low in animal fats.

'Western' characterizes only the international distribution of diseases. It does not describe other features of disease distribution in populations. In particular, it does not predict changes in incidence other than the rise with the beginnings of industrialization.

Two main environmental changes during industrialization are in hygiene and in diet. Improvements in sanitation and housing lead to falls in mortality from infective diseases, especially among children². Increased intake of total calories, animal fat, fruit and fresh vegetables, vitamins and minerals leads to better growth and development of children and contributes to the decline of infective diseases such as tuberculosis.

Epidemiological findings suggest that many diseases in industrialized countries fall into one of three groups, which can be shown by data from England and Wales, where numbers and causes of all deaths have been recorded for 140 years. The first group, associated with affluence, includes gallstones, renal stones and cancers of the breast, ovary and prostate. The incidence of these diseases has risen steeply during this century, particularly in more prosperous areas. The second group, associated with poor living standards, includes chronic bronchitis, stroke, stomach cancer and rheumatic heart disease. The incidence of these diseases has fallen during the past 50 years as living standards have improved. These diseases are commonest in the least affluent areas and in people with the lowest incomes. The third group is at different times associated with both affluence and poverty, and includes ischaemic heart disease, appendicitis and duodenal ulcer. The incidence of these diseases rose in the early part of this century, when they were more common among the rich. Subsequently the incidence fell and they are now commoner among poorer people.

The annual death rates from this last group of diseases in England and Wales³

are shown in Fig. 1. Together with those in the first group, they can be termed 'Western' diseases. Why have they declined when the environmental changes of industrialization have persisted?

Poliomyelitis was rare in Britain before this century but, as in other countries, it began to appear as hygienic and general living standards improved. This is in contrast to all the other common infective diseases, which were declining at this time. There was a sharp rise in notifications of poliomyelitis after the Second World War which persisted until the introduction of large-scale immunization (Fig. 1). It is now known that the rise of poliomyelitis resulted from the increasing vulnerability of the central nervous system to poliovirus infection with increasing age. As hygiene, sanitation and housing improved, the proportion of children escaping infection during the relatively safe period of infancy rose, and the number of cases of paralytic disease at later ages therefore rose in parallel.

The outbreaks of appendicitis which have accompanied industrialization in many parts of the world can similarly be explained as an age-dependent consequence of infection⁴. In England and Wales, death rates from appendicitis increased abruptly and steeply from around 1900, and then fell progressively from the 1930s onwards (Fig. 1). There is strong evidence that this trend reflects

changes in incidence of the disease, and similar trends were recorded in other European countries and in the United States. The explanation of these trends is thought to lie in the reduced levels of enteric infection in young children brought about by better hygiene, making them liable to develop appendicitis in response to infections at a later age. With continued improvements in hygiene, exposure to infection throughout childhood and early adult life became more uncommon. Because the appendix seems to be less vulnerable to infection after about 30 years of age, appendicitis declined.

Death rates from duodenal and stomach ulcers rose and fell in a similar way. The distribution of duodenal ulcer with social class was similar to appendicitis; while it was increasing it was more common among the rich, but as it declined it became commoner in poorer people. Recent findings of *Campylobacter*-like organisms in peptic ulcers suggest that the disease is spread by an infective agent⁵. The time trends point to age-dependent consequences of infection.

The trends in deaths from thyrotoxicosis (Fig. 1) show another process by which a rise in disease can be followed by a fall. Death rates rose to a peak in the 1930s and thereafter declined. Most deaths from this disease occur in the elderly, among whom toxic multinodular goitre is the usual cause of thyrotoxicosis. Analyses

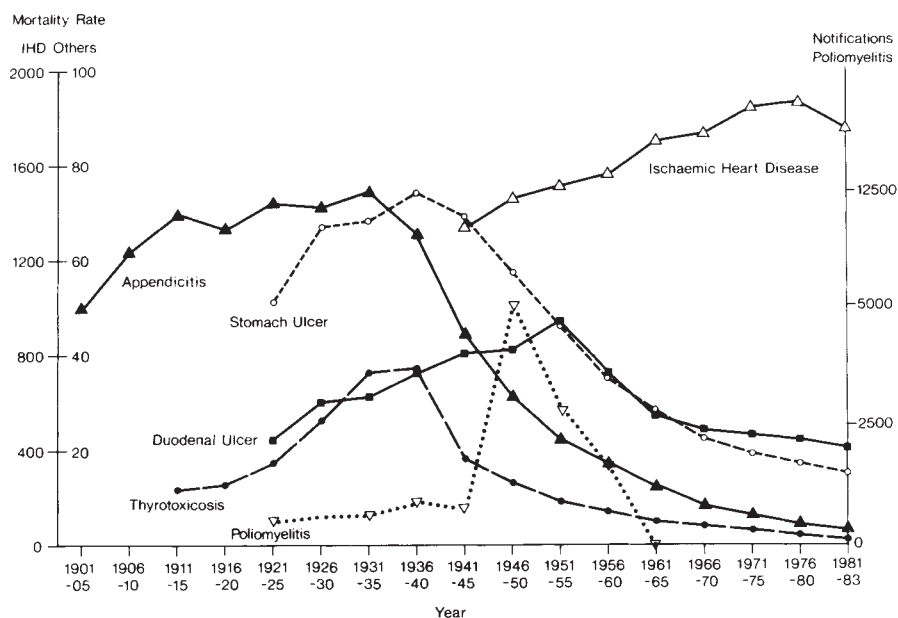


FIG. 1 Average annual mortality from selected diseases in England and Wales from 1901, and numbers of notifications of poliomyelitis in five-year periods³.

of age-specific rates in successive generations⁶ show that rates rose progressively in people born after 1836 and reached a peak in those born between 1871 and 1886. This is shown in Fig. 2, which gives rates in each generation according to their year of birth.

Figure 2 also shows the progressive increase in dietary iodine in Britain during this century, as a consequence of diversification of diet and availability of iodine in many foods including fish, meat and milk⁷. Iodine deficiency during childhood was widespread among people born in Britain in the 1800s. Successive generations, however, were exposed to more iodine in adult life. There is evidence that people who are iodine deficient in youth are less able to adapt to increased iodine intake in later life and tend to develop thyrotoxicosis⁸. This would explain the rise in deaths from thyrotoxicosis in the early part of this century (Fig. 1). Successive generations born after 1880 were exposed to more iodine in childhood, which would have lessened their susceptibility to iodine in adult life; accordingly, thyrotoxicosis mortality fell from around 1940 (Fig. 1). This explanation of the time trends accounts for the apparent paradox that toxic nodular goitre is now common only in those areas of Britain where iodine deficiency used to be prevalent.

The essential process thought to underlie the trends of occurrence of toxic nodular goitre is a response to an environmental influence during early life, which has a critical effect on the ability to adapt to subsequent exposure. The same process may determine trends in ischaemic heart disease. Death from ischaemic heart disease was not distinguished from that from other forms of heart disease before 1940, but other evidence reveals a steep rise in rates during the first part of the century. From 1940, death rates continued to rise progressively and reached a plateau in the 1970s (Fig. 1). Since 1980 there has been a small decline. Similar patterns occurred in the United States, Canada, Australia and New Zealand; initial rises were followed by substantial falls, of around one-quarter in the past 20 years in the United States⁹.

Although death from ischaemic heart disease was more common in wealthier people during its steep increase in Britain, it is now more common in less affluent areas and among people with lower incomes. This geographical distribution is the reverse of that for other diseases thought to be caused by the Western diet — gallstones, renal stones and breast cancer are all more common in more prosperous areas. The geographical distribution of ischaemic heart disease is closely related to that of poor child health and development, indicated by high infant and child mortality about 70 years ago¹⁰. The aetiology of this disease may therefore

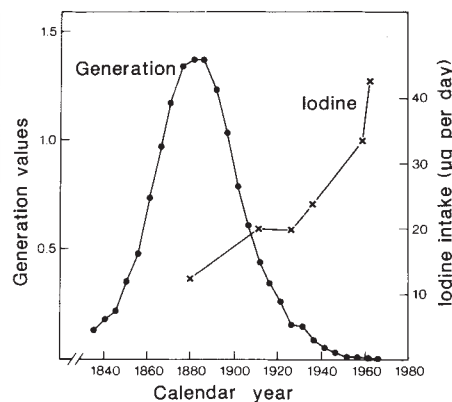


FIG. 2 Relative mortality from thyrotoxicosis in successive generations of women in England and Wales according to year of birth, together with estimated per capita daily iodine intake from milk, meat and fish⁷.

depend on two groups of environmental causes: one associated with affluence and likely to be mediated through a high-energy, high-fat diet; the other acting during childhood and associated with poor living standards. The rise in the disease results from an increase in the first, its fall from reduction in the second.

Further evidence of the effect of this second group of environmental causes is the inverse relation between risk of ischaemic heart disease and height, which is largely determined by growth in childhood. Long-term 'programming' of lipid metabolism during infancy, in response to infant feeding, is a mechanism for which there is increasing evidence in experimental animals and which could be a link between childhood and ischaemic heart disease. Another link may be the known inverse relation between fetal growth and blood pressure in adult life¹¹.

Each of the diseases shown in Fig. 1 may rise and then fall in response to one environmental change that accompanies industrialization. Such diseases characterize the change from a mainly rural to a mainly industrial society. Responses to the childhood environment, including age-dependent consequences of infection and adaptation to diet during early life, may be important causative factors.

The search for causes of 'Western' diseases has concentrated on the adult environment. The importance of the childhood environment in determining responses throughout life may have been underestimated. Models of disease based on the effects of cigarette smoking, an influence in the adult environment which has been intensively studied, may have limited general application. Where differences in individuals' susceptibility to disease cannot be explained by differences in the adult environment, as is the case for ischaemic heart disease, they have often been attributed to genetic causes — especially if the disease has a familial tendency. Part of what is now regarded as the genetic contribution to ischaemic

heart disease may turn out to be the effect of the intra-uterine or early postnatal environment.

Critical adaptations during early life may determine optimal rates of change within populations. Appendicitis and duodenal ulcer became common at an early stage of improvements in hygiene. The size of epidemics of these diseases may depend on the speed with which hygiene improves throughout the population. The rise of appendicitis in Britain can be linked to the introduction of domestic hot-water systems. The introduction of piped water supplies was spread over more than half a century, piped water not reaching some rural areas until after the Second World War. Swifter execution of public-health reforms begun in the nineteenth century might have reduced the incidence of this disease.

By contrast, critical responses to nutrition in childhood, such as those that occur in toxic nodular goitre and that are suspected in ischaemic heart disease, may limit the extent of dietary change that a generation can be exposed to without adverse effects. It follows that improvements in living standards during early industrialization should be directed at children. Although the industrial revolution in Britain brought high wages to adults, children continued to grow up undernourished, in large families, and in poor, overcrowded homes.

Steep increases in the incidence of 'Western' diseases regularly follow industrialization in the developing world. Among people of Chinese origin, improvements in hygiene may occur without changes in the traditional diet. Elsewhere, as in the slums of many cities in the developing world, diet changes but poor hygiene persists. If more was known about the processes by which the environment in early life influences adult health, the hygienic and nutritional benefits which will accompany industrial development might be maximized, and the rise in incidence of 'Western' disease minimized. □

David J. P. Barker is Director of the MRC Environmental Epidemiology Unit, Southampton General Hospital, Southampton SO9 4XY, UK.

1. Trowell, H.C. & Burkitt, D.P. *Western Diseases: Their Emergence and Prevention* (Arnold, London, 1981).
2. McKeown, T. & Lowe, C.R. *An Introduction to Social Medicine* (Blackwell Scientific, Oxford, 1974).
3. *The Registrar General's Statistical Review of England and Wales Part 1* (HMSO, London, 1901 et seq).
4. Barker, D.J.P., Osmond, C., Golding, J. & Wadsworth, M.E.J. *Br. Med. J.* **296**, 956–958 (1988).
5. Marshall, B., McGee, D., Rogers, P. & Clancy, R. *Med. J. Aust.* **142**, 439–444 (1985).
6. Phillips, D.I.W., Barker, D.J.P., Winter, P.D. & Osmond, C. *J. Epidemiol. Community Health* **37**, 305–309 (1983).
7. Greaves, J.P. & Hollingsworth, D.F. *World Rev. Nutr. Diet* **6**, 34–89 (1966).
8. Barker, D.J.P. & Phillips, D.I.W. *Lancet* **ii**, 567–570 (1984).
9. Pisa, Z. & Uemura, K. *World Hlth Stat. Q.* **35**, 11–47 (1982).
10. Barker, D.J.P. & Osmond, C. *Lancet* **i**, 1077–1081 (1986).
11. Barker, D.J.P., Osmond, C., Golding, J., Kuh, D. & Wadsworth, M.E.J. *Br. Med. J.* **298**, 564–567 (1989).

Heat conduction is a can of worms

Heat conduction is usually regarded as a phenomenon in which energy diffuses away from regions of high temperature, but diffusion may not always be a sufficient explanation.

WHETHER it is pleasing or otherwise to be told that a familiar principle is also a pack of lies is a matter of temperament. Some people are natural iconoclasts, others wish that at least some of the basic verities were eternal. But there are good reasons why what is known as Fourier's law (of heat conduction), which many people are still taught at high school, should be regarded as the crudest approximation to the truth. The interesting, but perplexing, question is that of what should take its place.

Most simply put, Fourier's law relates the flux of heat at any point in, say, a one-dimensional solid such as a metal bar to the temperature gradient at that same point. The heat flux and temperature gradients are taken as proportional to each other, with a constant of proportionality called thermal conductivity and with a minus sign (to allow that heat flows against the temperature gradient). The relationship is that of the diffusion equation, describing the diffusion of a chemical in a liquid.

High-school problem-solvers know how this can be used to calculate the change of temperature with time at any point in, say, a conducting bar. The rate of change with time of the temperature at any point, determined by the rate at which energy accumulates there, is evidently proportional to the rate of change with distance of the heat flux, with a constant which is the reciprocal of the specific heat and with another minus sign. The outcome is an equation relating the first time-derivative of the temperature and the second distance-derivative of the temperature. Given the distribution of temperature at the outset, it is then possible to calculate the temperature everywhere at all later times.

One might, for example, start with a metal bar at room temperature for half its length and at some higher temperature for the other half and follow the approach to equilibrium, when the temperature along the bar will be everywhere the same. But there is one obvious respect in which this result flies in the face of common sense: all solutions of these equations entail that the temperature everywhere along the bar will begin to change from the outset, however long the bar.

The implication is that the influence of the hotter half of the bar propagates instantaneously even to the most distant parts of the cooler half, which is unphysical. It is not merely that relativity disallows influences propagating faster than the speed of light, but that heat must get

from one end of the bar to the other by some physical process plainly overlooked by Fourier's law.

This is the starting point for a review earlier this year by D. D. Joseph of the University of Minnesota and Luigi Preziosi of the University of Naples (*Rev. Mod. Phys.* **61**, 41; 1989). Their objective is to define the circumstances in which heat can be propagated by means of waves, which has been observed in liquid helium and in some dielectric crystals at low temperatures. But their review inevitably begins with a discussion of the mechanisms of heat conduction and turns out to be a stimulating history of attempts to improve on Fourier's law, with the loose ends left dangling for all to see.

That there should be loose ends is not surprising; the mechanisms of heat transport are many and different. In a gas, energy is physically carried by molecules and transferred whenever they collide. In solids, on the other hand, lattice vibrations are a universal means of heat transport, but free electrons also play an important part in metals.

In all cases, the physical transport of heat from one place to another requires that some physical interaction should release energy at the destination — in a gas, the collision of an energetic molecule or, in a solid, the conversion of a quantum of energy in one vibrational mode into one or more quanta tied to other modes (which is possible only because of the anharmonicity of the vibrations of real crystals). Improvements on Fourier's law must take account of these processes, which involve microscopic energy conversion with characteristic times and distances (the time between collisions and the mean free path in a gas, for example).

Nobody will be surprised that James Clerk Maxwell is credited with the first important contribution to the argument, in 1867 (*Phil. Trans. Roy. Soc. Lond.* **157**, 49; 1867). Following logic and his instincts, Maxwell (working with his kinetic theory) concluded that the spatial gradient of temperature is not simply related to the heat flux, but to a combination of that and a quantity proportional to the time-rate of change of the heat flux. The constant of proportionality must evidently have the dimensions of time; numerically, it is a 'relaxation time' related (in Maxwell's case) to the interval between molecular collisions.

Joseph and Preziosi observe that Max-

well, having arrived at this conclusion, promptly discarded the extra term he had imported into the equation of Fourier's law on the grounds that the relaxation time must be negligible. In the event, his equation was rediscovered only in 1948, by C. Cattaneo. Mathematically, Fourier's law in its simple form boils down to a differential equation for the temperature (as a function of position and time) which is strictly equivalent to the diffusion equation (which is why the ratio of the thermal conductivity and the specific heat of a material is called its thermal diffusivity). Maxwell's formulation, and that of Cattaneo, are by contrast equivalent to the partial differential equation known as the 'telegraph equation', which is itself a modification of the more familiar wave equation and whose solutions are waves which are attenuated as they propagate. So why do we not more often encounter phenomena in which heat seems to propagate as waves, not by diffusion?

As early as 1917, Nernst was arguing intuitively that, at low temperature, propagating heat should have inertia of a kind; a pulse of internal energy arriving at some point in one direction might be expected to continue in the same direction in the absence of the influences that, at high temperature, would randomize the internal motions. That, said Nernst, should make wave propagation possible. In 1941, Landau developed the two-fluid theory of liquid helium II and predicted the existence of heat waves, recognizable as temperature oscillations, which he called 'second sound'. The prediction was confirmed in 1944, when Peshkov measured the speed of second sound in helium II. Peshkov also predicted wave-like heat propagation in low-temperature crystals, also now confirmed.

So why not room-temperature heat waves? Joseph and Preziosi take the view that the question cannot be decided until more is understood about the spread of relaxation times in the process of heat conduction. Meanwhile, they advocate a further modification of the equation corresponding to Fourier's law to take account of these refinements. Luckily, as their history of the field shows, all kinds of new techniques — molecular dynamics, for example — are now being brought to bear on the issue. And, for practical purposes, Maxwell's guess that the simple form of Fourier's law suffices for most purposes remains valid. **John Maddox**

All for one, one for all, that is our device

Jon Seger

SOCIAL insects have played a unique role in the development of the theory of natural selection, ever since Darwin noted¹ that sterile workers pose a "special difficulty, which at first appeared to me insuperable, and actually fatal to my whole theory". Darwin saw in outline how selection might be "applied to the family, as well as to the individual", an idea that eventually gave rise to the modern theory of inclusive fitness². Inclusive fitness is a general concept that applies to all situations in which relatives interact, but the motivation to develop it (and in particular, its connections with sex-ratio theory^{3,4}) came largely from social insects. Theories of group selection⁵ have also been inspired by insect sociality, and may soon receive new impetus from the same source.

On page 420 of this issue⁶, Rissing and colleagues show that unrelated females of the desert leaf-cutting ant *Acromyrmex versicolor* seem to participate voluntarily in an unequal division of risk and labour during the establishment of new colonies.

Several different patterns of colony formation are found among social insects of the order Hymenoptera, and most have evolved more than once⁷. In some ants and bees, for example, a fraction of the parent colony's worker force departs with a newly produced female reproductive (gyne) who becomes the principal egg-layer (queen) of the resulting daughter colony. But in most social Hymenoptera, each gyne departs alone and attempts to establish a new nest and to raise a first

generation of workers without any help at all. In either case a colony has a single queen, and the relatedness of nestmates tends to be high.

The new colony may also be established by several or many gynes, without an initial worker force. This pattern is common among social wasps. Dominance hierarchies form among the cofoundresses, such that one or several high-ranking individuals monopolize reproduction, and subordinates act as workers. The average relatedness of nestmates varies widely, but cofoundresses seem to be significantly related in most species⁸.

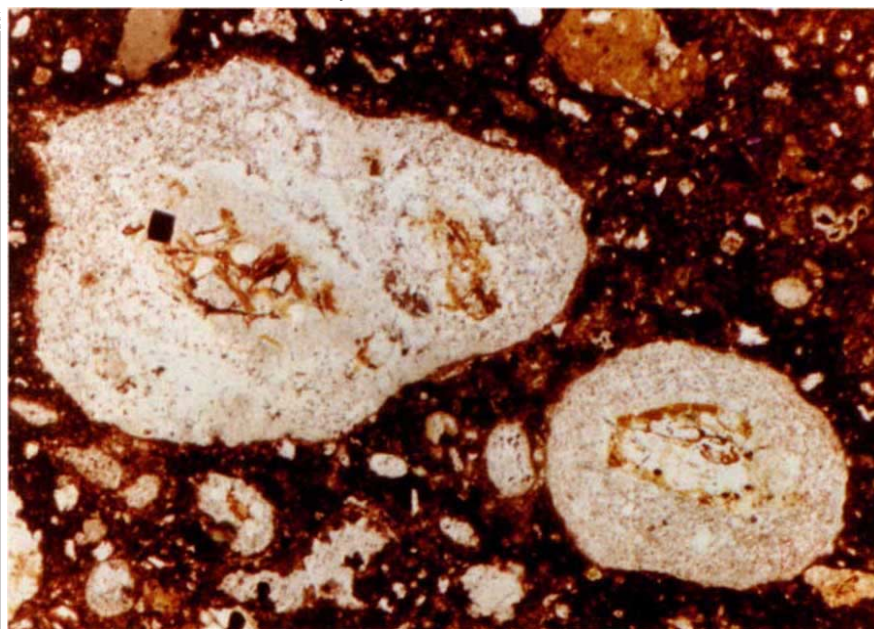
Multiple foundresses also occur in some species of ants: by analogy to social wasps, one would expect them to be relatives. In 1910, the great myrmecologist William Morton Wheeler⁹ assumed "without doubt" that two cofounding females "were sisters that had accidentally met . . . and had renewed the friendly relations in which they had lived before taking their nuptial flight". Only in the past decade has it been learned that, in ants, cofoundresses are usually unrelated. The first evidence came from behavioural studies showing that gynes readily team up with each other independent of the distance between their natal colonies (and thus presumably independent of their relatedness)^{10,11}. Recent genetic studies of relatedness within colonies of *Solenopsis invicta*¹², *Veromessor pergandei*¹³ and *A. versicolor*¹³ show that cofoundresses are indeed no more closely related than random individuals drawn from the population at large.

In the laboratory, gynes have survived better in groups than alone¹⁰, and foundress groups of moderate size have produced more first-generation workers, more quickly, than have smaller or larger groups^{10,11}. Species that frequently found nests cooperatively often engage in reciprocal brood raiding, with small colonies tending to be eliminated by their larger neighbours^{10,11}. Thus at high population densities, cooperative founding may greatly increase the probability that a given nest will survive long enough to produce reproductives.

An alliance may improve survival, but given survival, it dilutes the expected reproductive success of each foundress. Conflict between foundresses or between workers and foundresses often begins soon after the emergence of the first workers^{10,11,14}, and such conflict often leads to the expulsion or killing of one or more foundresses; in many species with multiple foundresses this process goes to completion and the mature colony is monogynous¹¹.

In most ant species, foundresses raise the first workers on secretions derived metabolically from flight muscle and fat⁷, and they never leave the nest. *A. versicolor* is unusual in that one found-

Life imitates (experimental) art



D. K. Bailey

THE white globules in this micrograph are chilled droplets of a carbonate melt, set in a groundmass of carbonate-rich volcanic ash. There has been much controversy as to whether such 'carbonatite' magma can form directly by partial melting of the mantle, or whether a two-stage process, involving modification of a mantle melt at shallower, crustal levels, is required. Recently, Margaret Wallace and David Green (*Nature* 335, 343–346; 1988; see also the accompanying News and Views article by J. Gittins) reported the experimental formation of a carbonatite melt by melting a mantle peridotite composition at mantle pressures and temperatures, lending support to the idea that carbonatites can be primary melts from the mantle. Now Ken Bailey on page 415 of this issue reports the existence of a close natural analogue to this experimental melt — the droplets pictured above, from the Rufunsa volcanoes in Zambia. Although these carbonatites have been known for some time, the similarity of their composition to that produced in the experiments, together with the discovery by Bailey of inclusions of mantle-type chromium-rich spinel (such as the black diamond-shaped grain in the larger droplet) now provides evidence that they are the long-searched-for primary carbonate melts. Field of view is 3.4 mm wide.

Laura Garwin

TOPOLOGY

Lowering the volume

Ian Stewart

ress forages outside the nest during the interval between nest establishment and the emergence of the first workers⁶. Foraging is energetically costly and it exposes the forager to risk of predation. One might therefore expect the forager to be a subordinate foundress who was forced into the role by her dominant nestmates¹⁵, but Rissing and colleagues⁶ find no evidence of conflict within foundress associations studied in the laboratory, either before or after emergence of the first workers.

Why should one foundress (the forager) perform such heroic service on behalf of nestmates to whom she is not related? The answer suggested by Rissing and colleagues is that efficiency matters above all else, in the race to establish a colony that survives the brood-raiding that will begin as soon as any neighbouring colony produces a worker force. Like her sedentary nestmates, the forager will have no chance to reproduce in the distant future if her colony succumbs early on. If the fates of cofoundresses are very tightly bound together then the appearance of selflessness may be only an illusion.

This argument seems to work best for the critical early history of the colony; it loses force as the colony matures and achieves greater security. When the time comes to produce reproductives, each queen would presumably wish the others gone. Rissing and colleagues observed 1.5 years of peaceful coexistence in their laboratory colonies⁶, which may well be a record for cofounding ant species¹¹. The story will become even more remarkable if this solidarity among unrelated queens remains unshaken throughout their lives. □

Jon Seger is in the Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA.

ONE of the central aims of topology is to classify all possible manifolds up to topological equivalence. A manifold is a multi-dimensional analogue of a surface, and two manifolds are topologically equivalent if one can be deformed continuously into the other. The time-honoured method for classifying manifolds is to construct topological invariants: computable numerical or algebraic data that are the same for topologically equivalent manifolds but that distinguish between topologically inequivalent ones. For example, the number of holes in an ordinary two-dimensional surface is such an invariant. S.V. Matveev and A.T. Fomenko (*Usp. mat. Nauk.* **43**, 5–22; 1988; *Russ. math. Surv.* **43**, 3–24; 1988) have now made several fundamental advances concerning one of the most unusual and surprising invariants, the volume of a hyperbolic manifold.

The usual notion of volume is changed if the object concerned is stretched or bent, but this particular version of volume is extraordinarily 'rigid', and is not changed by continuous deformations. In addition, Matveev and Fomenko disprove a conjecture of Thurston (*Bull. Am. math. Soc.* **6**, 357–381; 1982) about the hyperbolic three-dimensional manifold (3-manifold) of minimal volume. Moreover, their methods have unexpected connections with fundamental questions in dynamics.

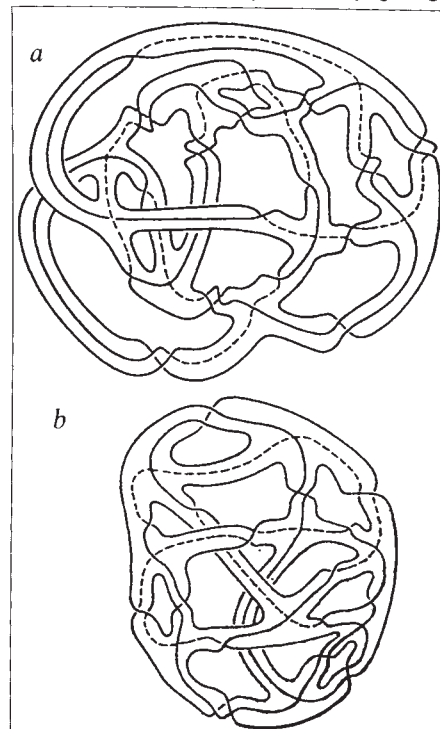
Over the past century, almost all of the objectives in the programme of finding invariants for manifolds have been accomplished for cases in which the dimension is either two, or greater than or equal to four. The remaining case, that of 3-manifolds, has proved far more subtle and intractable. The most encouraging new direction is embodied in recent conjectures of Thurston to the effect that every 3-manifold can be cut into pieces, each of which has a natural geometric structure (see my News and Views article in *Nature* **328**, 16; 1987).

This conjecture is at first quite startling, because geometry involves rigid concepts such as lengths and angles, which are destroyed by topological transformations. But the result has long been known to be true (and languished, unappreciated) in two dimensions. For two-dimensional surfaces there are three fundamentally different kinds of geometry: euclidean geometry; elliptic geometry, in which no parallel lines exist; and, in many ways the most interesting, hyperbolic geometry, in which there are infinitely many distinct lines through a given point and parallel to a given line.

Elliptic and hyperbolic geometry are the two possible types of non-euclidean

geometry in two dimensions. Elliptic geometry can be realized on the surface of a sphere, which has constant positive curvature; and hyperbolic geometry is valid on a surface of constant negative curvature. Euclidean geometry comes in between: it holds for flat spaces, of zero curvature.

A surface without boundary, such as a sphere or multi-holed torus, can be equipped with a unique geometric structure. For example, as already mentioned, the sphere carries an elliptic geometric structure. The standard torus, with one hole, can be obtained by abstractly 'gluing'



Combinatorial data defining *a*, Thurston's conjectured minimal-volume 3-manifold; and *b*, the new record holder identified by Matveev and Fomenko.

together the edges of a square. Because the square is contained in a flat plane, and no distortion of distances or angles occurs when its edges are 'glued', the torus carries the same euclidean geometric structure as its parent square. It turns out that all other surfaces, that is, tori with two or more holes, can be equipped with a hyperbolic structure.

Except in the euclidean case, the geometric structure has a natural length scale. Two spheres, for example, differ only in their radius. If the curvature (the reciprocal of the radius) is normalized to 1, then the radius must be 1. This fixes the length scale. The area must then be 4π ; and this number, so determined, is a topological invariant. Any surface, if it can be

1. Darwin, C. *On the Origin of Species By Means of Natural Selection* (Murray, London, 1859).
2. Hamilton, W.D. *J. theor. Biol.* **7**, 1–52 (1964).
3. Trivers, R.L. & Hare, H.H. *Science* **191**, 249–263 (1976).
4. Nonacs, P. *Q. Rev. Biol.* **61**, 1–21 (1986).
5. Wilson, D.S. *A. Rev. Ecol. Syst.* **14**, 159–187 (1983).
6. Rissing, S.W., Pollock, G.B., Higgins, M.R., Hagen, R.H. & Smith, D.R. *Nature* **338**, 420–422 (1989).
7. Wilson, E.O. *The Insect Societies* (Harvard University Press, 1971).
8. Queller, D.C., Strassmann, J.E. & Hughes, C.R. *Science* **242**, 1155–1157 (1988).
9. Wheeler, W.M. *Ants: Their Structure, Development and Behavior* (Columbia University Press, New York, 1910).
10. Bartz, S.H. & Holldobler, B. *Behav. Ecol. Sociobiol.* **10**, 137–147 (1982).
11. Rissing, S.W. & Pollock, G.B. in *Interindividual Behavioral Variability in Social Insects* (ed. Jeanne, R.L.) 179–222 (Westview, Boulder, 1988).
12. Ross, K.G. & Fletcher, D.J.C. *Behav. Ecol. Sociobiol.* **17**, 349–356 (1985).
13. Hagen, R.H., Smith, D.R. & Rissing, S.W. *Psyche* (in the press).
14. Carlin, N.F. in *Interindividual Behavioral Variability in Social Insects* (ed. Jeanne, R.L.) 147–177 (Westview, Boulder, 1988).
15. West Eberhard, M.J. in *Natural Selection and Social Behavior* (eds Alexander, R.D. & Tinkle, D.W.) 3–17 (Chiron, New York, 1981).



100 years ago

THE GEOGRAPHICAL RESULTS OF MR. STANLEY'S EXPEDITION

It is evident from Mr. Stanley's stirring letters, which during the past week have cast all other topics into the shade, that pioneering in Africa is not yet at an end, and that the strange continent has not yielded up its vast wonder to knowledge.

Mr. Stanley's description of the character and extent of the Congo forest in his letter to Mr. Bruce is quite worth quoting:-

"Take a thick Scottish copse, dripping with rain; imagine this copse to be a mere undergrowth, nourished under the impenetrable shade of ancient trees, ranging from 100 to 180 feet high; briars and thorns abundant; lazy creeks meandering through the depths of the jungle, and sometimes a deep affluent of a great river. Imagine this forest

and jungle in all stages of decay and growth — old trees falling, leaning perilously over, fallen prostrate; ants and insects of all kinds, sizes, and colours murmuring around, monkeys and chimpanzees above, queer noises of birds and animals, crashes in the jungle as troops of elephants rush away; dwarfs with poisoned arrows securely hidden behind some buttress or in some dark recess; strong brown-bodied aborigines with terribly sharp spears, standing poised, still as dead stumps; rain pattering down on you every other day of the year; an impure atmosphere, with its dread consequences, fever and dysentery; gloom throughout the day, and darkness almost palpable throughout the night; and then, if you will imagine such a forest extending the entire distance from Plymouth to Peterhead, you will have a fair idea of some of the inconveniences endured by us from June 28 to December 5, 1887, and from June 1, 1888, to the present date, to continue again from the present date till about December 10, 1888, when I hope then to say a last farewell to the Congo forest."

From *Nature* **39**, 560; 11 April 1889.

given a geometric structure of curvature 1, must have the same area 4π as the sphere — and indeed in this case must be equivalent to a sphere. In a similar way negative curvature can be normalized to -1 , and this also fixes a length scale. Therefore surfaces with a hyperbolic geometric structure also have uniquely defined areas.

For surfaces, far simpler invariants (such as the number of holes) achieve the desired classification, so this area has not featured prominently in most topologists' imaginations. But for 3-manifolds, analogously defined 'volumes' are turning out to be a much richer concept. As Thurston observed, there are eight distinct types of geometry in three dimensions. Three are euclidean, elliptic and hyperbolic geometry; the other five are harder to describe. Seven of these eight types of geometry are well understood, and so are the manifolds that carry such structures. The difficult case is that of hyperbolic manifolds: 3-manifolds that can be equipped with a geometry having constant negative curvature. For such manifolds one can normalize the curvature to -1 , again obtaining a well-defined length scale; so the manifold has a uniquely determined volume. Moreover, this volume is a topological invariant — and a remarkably effective one.

Most invariants tend to be whole numbers. A simple example is the number of holes — clearly one cannot have half a hole. But the volume is a decimal. It is known that only a finite number of hyperbolic 3-manifolds have dimensions less than any given value. In particular there must be a smallest value. The corresponding manifold is presumably of some importance — otherwise why should it be the smallest one? So, which is it? Thurston conjectured that a particular manifold has

this smallest possible volume, and that its value is 0.98139.

Matveev and Fomenko use a new approach, combining two very distinct ideas: the complexity of a 3-manifold, and connections with dynamical systems. Complexity measures how complicated a combinatorial description of the manifold must be; the dynamical connection is described below. The complexity of a 3-manifold, which is a whole number, has many pleasant properties. For example, only finitely many manifolds have given complexity. The number of distinct 3-manifolds of complexity 0, 1, 2, 3, 4, 5 and 6 is precisely 3, 2, 4, 7, 14, 31 and 74, respectively. None of these is hyperbolic — in fact the first hyperbolic manifolds occur at complexity 9, and one of these is Thurston's conjectured manifold of minimal volume.

Using this notion of complexity, Matveev and Fomenko set up a computer algorithm to search for hyperbolic 3-manifolds and compute their volumes. They studied about a thousand manifolds in all. Among their list they found a new manifold, also of complexity 9, with the forbidding name $(Q_1^3)_{8-2}$, whose volume 0.94272 is smaller than that conjectured by Thurston. Indeed it is the only manifold in their list having smaller volume than Thurston's manifold, and they in turn conjecture that they have now found the true minimal volume. Because 3-manifolds involve 'curved space' they cannot be drawn directly, but their curious and intricate geometry may be gathered from two pictures (see the figure) showing the so-called singular graph for Thurston's manifold and the new, low-volume record holder. The singular graph determines certain combinatorial data from which the manifold can be constructed.

Matveev and Fomenko's very extensive

list of manifolds and volumes provides a wealth of valuable data against which to test conjectures. In particular their results cast doubt on one part of Thurston's programme for studying 3-manifolds. He suggests dissecting manifolds into simplexes (non-euclidean tetrahedra) and analysing how these fit together. But it now seems that some features of the mathematics depend on the manner in which this dissection is done, an unexpected difficulty.

Finally, there is a curious and deep connection with dynamics, which is central to Matveev and Fomenko's theory. Every hamiltonian dynamical system (that is, one in which there is no friction, so that energy is conserved) has associated with it a phase space determined by the position and velocity variables needed to specify its state. For example a mass moving in space requires three position coordinates and three velocity coordinates, and hence has a six-dimensional phase space. The subset of phase space corresponding to a fixed value of the energy is a manifold one dimension smaller — in this case 5.

For a simpler example, think of a simple harmonic oscillator (or idealized pendulum). This has one position coordinate and one velocity coordinate, giving a two-dimensional phase space. The constant-energy manifolds are concentric circles. But circles have positive curvature, hence elliptic geometry. Hyperbolic energy surfaces do not occur in two-dimensional phase spaces.

Matveev and Fomenko prove that something similar happens when there are two position coordinates and two velocity coordinates. The phase space then has dimension 4, and every constant-energy manifold has dimension 3. Thus a special class of 3-manifolds arises out of dynamics. Which? Under one further technical assumption, integrability, these manifolds can be characterized. They are precisely those 3-manifolds which can be cut into pieces, each having a geometric structure that is non-hyperbolic. Any of the other seven possible geometries may occur: hyperbolic geometry alone is forbidden.

Some of these ideas may sound outlandish, but they derive from long-standing themes in mainstream mathematics. What is new is the manner in which they have been brought together. Volume, a rigid geometric concept, now applies in a topological context; it is related to complexity; and hamiltonian dynamics somehow conspires to forbid even a tiny part of a constant-energy surface having constant negative curvature. The story is a remarkable and important affirmation of the power of the new geometric approaches to topology and dynamical systems. $\square$

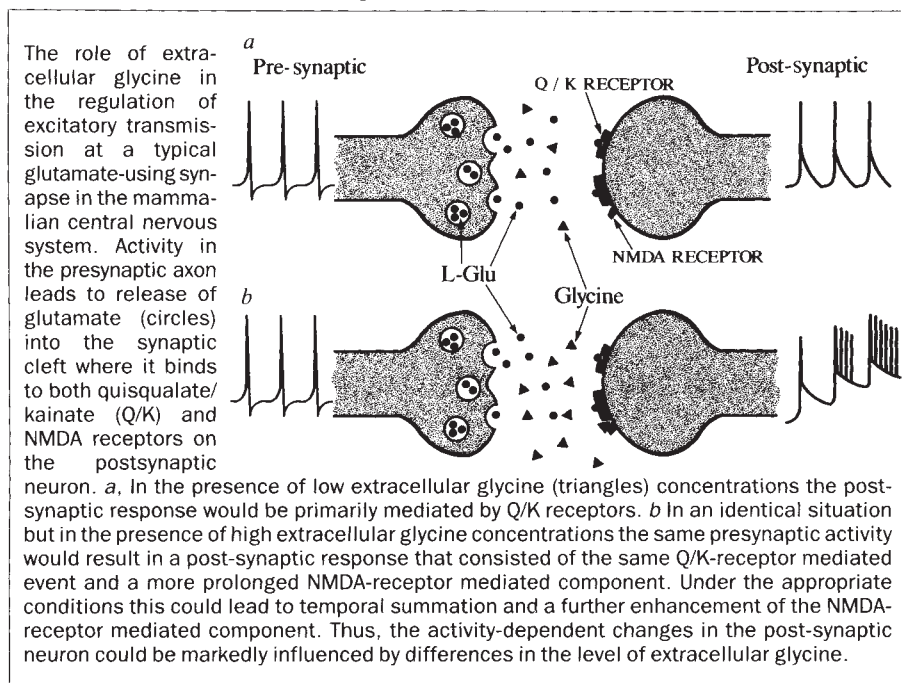
Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Glycine maintains excitement

Alan C. Foster and John A. Kemp

THE explosion of interest in the NMDA (*N*-methyl-D-aspartate) receptor, one receptor subtype for the excitatory neurotransmitter L-glutamate, has generated several new concepts of transmitter function. Thus, it is now known that NMDA-receptor activation is conditional on a unique voltage-dependent regulation of channel conductance by magnesium^{1,2}, and that the neuronal response consists of both short-term membrane conductance and long-term biochemical changes, the

concentrations of glycine increase the frequency of NMDA-receptor channel opening in a strychnine-insensitive manner, and suggested a mechanism involving allosteric regulation of the receptor complex through a distinct glycine binding site. This idea is supported by the existence of strychnine-insensitive ³H-glycine-binding sites which have an anatomical distribution identical to that of the NMDA receptor⁷. It seems likely that the glycine-binding site is an integral



latter caused by elevated intracellular calcium concentrations³. Furthermore, receptor activation involves not only the binding of an agonist, such as L-glutamate or NMDA, but also an allosteric facilitation by glycine binding to a separate site within the receptor complex⁴. Two papers in this issue provide further insights into this effect of glycine. On page 425, Mayer *et al.*⁵ present evidence that at least part of the enhancement of NMDA responses by glycine occurs through a previously unknown mechanism — an acceleration of recovery from desensitization. And on page 422, Thomson *et al.*⁶ demonstrate that glycine facilitates NMDA receptor-mediated neuronal excitation in slices of cerebral cortex, suggesting that glycine is involved in the tonic regulation of transmission.

The effects of glycine on the NMDA receptor are distinct from its long-established role as an inhibitory neurotransmitter in lower brain areas and spinal cord, mediated by a strychnine-sensitive chloride conductance. Johnson and Ascher¹ have shown that submicromolar

part of the receptor complex, because glycine facilitation is maintained in isolated membrane patches⁴ following solubilization of NMDA receptors⁸, and in NMDA receptors expressed in *Xenopus* oocytes following injection of rat brain messenger RNA⁹. The effects of glycine on the NMDA receptor have been compared with those of benzodiazepines on the GABA_A receptor, but the magnitude of the enhancement of NMDA responses by glycine is much greater than that seen with benzodiazepines on GABA responses. Indeed, Kleckner and Dingledine¹⁰ claim that no NMDA-receptor response can be obtained in the absence of glycine, and propose that receptor activation requires the binding of both glycine and L-glutamate as 'co-agonists'.

The new findings of Mayer *et al.*⁵ are of direct relevance to this problem and add a new facet to the glycine story. They suggest that the main effect of glycine is to prevent desensitization, a decline in the receptor response which occurs during a prolonged agonist application of the NMDA receptor. Glycine does not block

the onset of desensitization, but accelerates the recovery of the receptor from its desensitized state.

This is the first time such a mechanism for the regulation of a ligand-gated ion channel has been proposed; it may explain the large enhancement of NMDA responses achieved by glycine. If this were the sole mechanism through which glycine exerts its effects, then the co-agonist idea would be negated. But patch-clamp studies with the selective glycine antagonist 7-chlorokynurenate, using a fast perfusion system similar to that of Mayer *et al.*, indicate that this compound can produce a complete block of NMDA responses¹¹. Thus, glycine binding could have two effects, one involved with the initiation of NMDA-receptor activation and the other to speed up recovery from desensitization.

Many questions surround the role of glycine in the regulation of NMDA receptors *in vivo*. NMDA responses obtained in slices of adult central nervous system *in vitro* are generally not enhanced by glycine (or its analogue D-serine), but can be blocked in a glycine-reversible manner by glycine antagonists such as kynurenate, 7-chlorokynurenate and HA-966¹¹⁻¹³. Thus, the facilitation of NMDA responses by glycine occurs in adult neurons, but under these experimental conditions extracellular glycine levels are sufficiently high to occupy all of the glycine-binding sites.

It has been suggested⁴ that this situation may be the case *in vivo*, as levels of glycine in the cerebrospinal fluid are high (greater than one micromolar). Thompson *et al.*⁶ report that glycine enhances NMDA responses in slices of cortex when applied in close proximity to the neuron under study in conditions of reduced perfusion, which are claimed to mimic more closely the situation *in vivo*. This agrees with recent studies *in vivo* which have shown a facilitation of NMDA responses by locally applied glycine or D-serine¹⁴⁻¹⁶.

Thompson *et al.* also demonstrate a

- Nowak, L. *et al.* *Nature* **307**, 462-465 (1984).
- Mayer, M.L. *et al.* *Nature* **309**, 261-263 (1984).
- MacDermott, A.B. *et al.* *Nature* **321**, 519-522 (1986).
- Johnson, J.W. & Ascher, P. *Nature* **325**, 529-531 (1987).
- Mayer, M.L. *et al.* *Nature* **338**, 425-427 (1989).
- Thomson, A.M. *et al.* *Nature* **338**, 422-424 (1989).
- Bristow, D.R. *et al.* *Eur. J. Pharmac.* **126**, 303-308 (1986).
- McKernan, R. *et al.* *J. Neurochem.* **52**, 777-785 (1989).
- Verdoorn, T.A. *et al.* *Science* **238**, 1114-1116 (1987).
- Kleckner, N.W. & Dingledine, R. *Science* **241**, 835-837 (1988).
- Kemp, J.A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 6547-6550 (1988).
- Watson, G.B. *et al.* *Neurosci. Res. Commun.* **2**, 169-174 (1988).
- Fletcher, E.J. & Lodge, D. *Eur. J. Pharmac.* **151**, 161-162 (1988).
- Larson, A.A. & Beitz, A.J. *J. Neurosci.* **8**, 3822-3826 (1988).
- Danysz, W. *et al.* *Brain Res.* **479**, 270-276 (1989).
- Salt, T. *Brain Res.* **481**, 403-406 (1989).
- Forsythe, I.D. *et al.* *J. Neurosci.* **8**, 3733-3741 (1988).
- Davies, J. & Watkins, J.C. *Brain Res.* **59**, 311-322 (1973).
- Kemp, J.A. & Sillito, A.M. *J. Physiol. Lond.* **323**, 377-391 (1982).

facilitation of synaptic responses by glycine. This was observed only in the fraction of synaptic responses which exhibited an NMDA receptor-mediated component under these experimental conditions. This effect of glycine is likely to be a widespread phenomenon, as similar observations have been made in hippocampal neurons in culture¹⁷ and in thalamic neurons *in vivo*¹⁶; and can be inferred at various synapses in the central nervous system from earlier studies using HA-966^{18,19} which is now known to be a selective glycine antagonist¹³. Thus, glycine facilitation of synaptic responses mediated by NMDA receptors may be a common regulatory mechanism at excitatory synapses.

Clearly, it is now of interest to define the processes that regulate extracellular glycine concentrations in regions of the brain where NMDA receptors play important roles. Alterations of these systems would lead to profound changes in

neuronal activity mediated by NMDA receptors (see figure). The possibility that endogenous glycine antagonists (such as kynurenate) may be involved in regulating neuronal function should also be considered. If the main facilitatory effect of glycine is to accelerate recovery from desensitization of the NMDA receptor, it is likely that glycine will act in physiological situations where prolonged responses are required and repeated activation is necessary, such as induction of long-term potentiation. Conversely, a pathological excess of glycine could lead to prolonged excitation and epileptiform activity, contributing to the overactivation of NMDA receptors suspected to occur in seizure disorders and neurodegenerative disease. □

Alan C. Foster and John A. Kemp are at the Merck Sharp and Dohme Research Laboratories, Neuroscience Research Centre, Terlings Park, Eastwick Road, Harlow, Essex CM20 2QR, UK.

EXTREME ULTRAVIOLET ASTRONOMY

A new window on the Universe

P. Gondhalekar and C. Jordan

THERE are few spectral ranges in which observations of astrophysical sources have not been made. One of the few is the extreme ultraviolet region, which is virtually unexplored owing to earlier assumptions that the interstellar medium would be opaque at wavelengths between the hydrogen Lyman photoionization edge (91.2 nm) and about 30 nm. But during the past decade there has been increasing evidence that the absorbing interstellar material is not uniformly distributed. The growing optimism concerning the possibility of detecting this unexplored radiation from various sources was reflected at a recent meeting*. It became clear that observations in the extreme ultraviolet could provide crucial tests of current theories concerning objects as diverse as active galactic nuclei, cool dwarf stars and the jovian aurora. The growing confidence in, and enthusiasm for, the new astronomy was amply demonstrated by the presentation of many plans for suitable telescopes for future space missions.

The turning point for extreme-ultraviolet astronomy came in 1975 with the detection of emission from the white dwarf HZ43, by an instrument on the Apollo-Soyuz mission. More recently, several nearby hot white dwarfs were observed with the Voyager ultraviolet spectrometers (J.B. Holberg, University of Arizona). Soft X-ray spectra of some sources, including cool stars in the proximity of the Sun, were obtained with the X-ray satellite, EXOSAT (N.E. White,

ESTEC) which was sensitive to wavelengths up to about 40 nm.

Also, the case for observing Solar System objects in the extreme ultraviolet has already been established by spectra of the giant planets and their satellites, obtained from Voyager 1 and 2. For example, strong lines of oxygen and sulphur ions are observed from the Io plasma torus. Improved spectra extending to wavelengths below 60 nm should allow an improved understanding of the plasma processes causing the jovian auroral emission as ions precipitate from the magnetosphere (P.D. Feldman, Johns Hopkins University; W.M. Harris and J.T. Clarke, University of Michigan; S. Chakrabarti, University of California, Berkeley).

The local interstellar medium is of great interest because the distribution and density of the local gas determines the direction and distance to which extreme ultraviolet sources can be detected. The most accurate neutral-hydrogen column densities come from measurements of ultraviolet absorption in spectra of discrete stellar sources, but these do not give sufficient spatial coverage to establish the global distribution. All-sky surveys at 21 cm, the principal hydrogen radiowave emission line, do give full coverage but the spatial resolution is too coarse to measure the hydrogen opacity in a particular direction. (C. Heiles, University of California, Berkeley; D. York and P. Frisch, University of Chicago).

There is a diffuse, hot (10⁶ K) component of the interstellar medium which is an important source for ionizing helium, but

its location and how pervasive it is are questions still much debated (F. Bruhweiler and K.-P. Cheng, Catholic University; P. Jackobsen, ESA). The overall picture that is emerging is that the local region is relatively highly ionized and thus has low opacity, rather than being devoid of material; there are higher neutral-hydrogen mass column densities at larger distances in the galactic plane, particularly towards the Galactic Centre. Large molecular-cloud complexes, such as those in Taurus and Auriga, add regions of high absorption but there are galactic longitudes (around 230°) where absorption is lower than average.

It seems unlikely that any small 'windows' exist through which extra-galactic sources might be viewed at the longer wavelengths (Heiles), but there is a good chance that active galactic nuclei and quasars could be observed up to wavelengths of 20 nm (H. Marshall, University of California, Berkeley). The general advice for observing to great distances is to observe out of the galactic plane, as the colder, more dense medium has a smaller scale height. There is sufficient uncertainty in the distribution of the hot and cold components however, to guarantee that extreme-ultraviolet surveys will add considerably to our knowledge and understanding of the interstellar medium.

There is little doubt that the local interstellar medium will be transparent to extreme-ultraviolet radiation from many types of stars. Existing observations of X-ray emission from main-sequence B (hot, blue-white) stars lead to models which predict the presence of two component, wind-driven shocks with temperatures in the range 10⁵–10⁷ K. New spectra could further test these theories (J. Cassinelli, University of California, Berkeley). The solar extreme-ultraviolet spectrum, which gives a good indication of what might be expected from cool main-sequence stars, has been extensively observed at high spectral and spatial resolution, and many spectroscopic diagnostic techniques for measuring electron densities and temperatures have been tested (G.A. Doschek, US Naval Research Laboratory).

Similar spectra of nearby cool stars would allow emission lines formed in the temperature range 2 × 10⁴–10⁶ K to be observed, leading to improved models of the structure and energy balance in stellar coronae (C. J.). At present, only average temperatures and densities can be found for stellar coronae, using broad-band X-ray measurements. Spectra and observations of the temporal variation of lines formed at different temperatures would allow more precise determination of coronal conditions and of the contribution made by stellar active regions. It is important to improve measurements of coronal densities and temperatures in order to

*Berkeley Colloquium on Extreme Ultraviolet Astronomy, Berkeley, California, 19–20 January 1989.

understand known correlations between coronal X-ray fluxes, chromospheric line fluxes and stellar rotation rates. These must be controlled ultimately by the non-thermal heating and the energy loss processes. A wide range of more exotic stellar objects, including symbiotic stars, AM Herculis binary stars, and cataclysmic variables, should also provide exciting

UBIQUITIN

A marriage of convenience or necessity?

Jonathan R. Warner

UBIQUITIN, a 76-amino-acid protein, is, as its name implies, ubiquitous in eukaryotic cells (see ref. 1 for a review). A complex enzymatic system attaches it to proteins, usually by peptide bond formation with the ϵ -amino of a lysine, either singly or in branched multiple ubiquitin structures. Ubiquitin can be a signal for the protein-degradation system of cells, much as a blaze on a tree serves as a signal for the logger. But in others its embrace is less lethal: it is reversibly attached to the histone H2A without causing its degradation². Its function here is unknown.

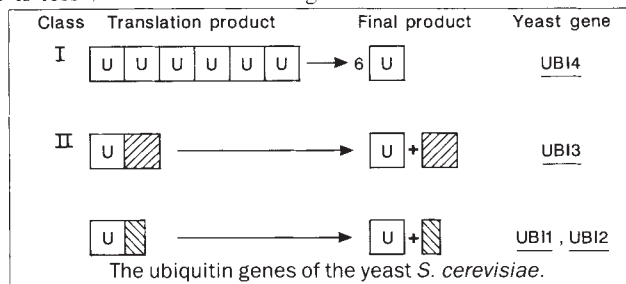
Ubiquitin is remarkably conserved through evolution, with a similarity of 96 per cent between yeast³ and human⁴, and between yeast and *Trypanosoma cruzi*⁵. Even more remarkable is the conservation in the arrangement of the genes encoding ubiquitin, which are found in two basic forms in most organisms (see figure). Class I is a polyubiquitin gene which encodes a polypeptide of up to 100 uninterrupted, tandemly repeated ubiquitins presumably released by proteolysis at the Gly-Met peptide bond that joins the repeats. Class II is a fusion between a single ubiquitin and one of two other sequences, of either 52 or 76–80 predominantly basic amino acids, again remarkably conserved through evolution. In this issue these basic sequences are identified, in the yeast *Saccharomyces cerevisiae* by Finley *et al.*⁶ on page 394, and in mammals by Redman and Rechsteiner⁷ on page 438, as members of another common group, ribosomal proteins. Finley *et al.*⁶ fused antigenic tags to the *UBI1* and the *UBI3* gene products of yeast and find the 52-amino-acid *UBI1* tail in the 60S ribosomal subunits and the 76-amino-acid *UBI3* tail in the 40S subunits. A protein with the amino-terminal sequence of the *UBI3* extension had previously been identified as ribosomal protein S37 (ref. 8). By immunoblotting with antibodies made against a peptide within the mammalian 80-amino-acid extension, Redman *et al.*⁷ find the same protein in the 40S subunits

new data (A.K. Dupree, S.J. Kenyon and J.C. Raymond, Harvard-Smithsonian Center for Astrophysics; D.Q. Lamb, University of Chicago). □

P. Gondhalekar is at the Rutherford Appleton Laboratory, Didcot OX11 0QX, and C. Jordan is in the Department of Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK.

and identify it as mammalian ribosomal protein S27a. Although less conserved through evolution than ubiquitin, yeast S37 and mammalian S27a are similar. Protein sequence data indicate that the ribosomal proteins from both organisms have been cleaved from the ubiquitin moiety without the loss of any amino acids.

By deletion analysis, Finley *et al.*⁶ show that the 52-amino-acid ribosomal protein is essential for cell growth, as are all other



eukaryotic ribosomal proteins they tested. It is therefore surprising that cells without the gene encoding ribosomal protein S37 do grow, although poorly. These cells have a severe deficiency of 40S ribosomal subunits, apparently because of a defect in the final processing step(s) in the formation of 18S ribosomal RNA. Does the lack of S37 influence only ribosome assembly, or does it also have some effect on the rate or accuracy of translation?

It is a remarkable finding that the sole source of two ribosomal proteins is derived from ubiquitin fusion genes. But it is truly extraordinary that this arrangement has persisted throughout eukaryotic evolution, even into the freakish trypanosomes. Is ubiquitin dependent on ribosomal proteins or vice versa? Or is there a mutual symbiosis?

Introns are rare in the genes of *S. cerevisiae*, whereas most of the genes encoding ribosomal proteins have a single intron. Although it is not yet clear what selective pressure has maintained introns in the ribosomal protein genes, it is interesting that the members of the duplicate pair, *UBI1* and *UBI2*, each have an intron in the ubiquitin-coding portion of the gene.

Another conserved region of ubiquitin

genes is the heat-shock promoter element^{3,5}. In several organisms, including yeast, the polyubiquitin genes are responsive to stress, but the fusion genes are constitutively expressed. In the case of the yeast genes, this is presumably because of the ribosomal protein enhancer elements, UAS_{mpg}, found upstream of the fusion genes³ and almost all ribosomal-protein genes^{9,10}. The fusion genes provide the main source of ubiquitin during log growth, when ribosome synthesis is active. The polyubiquitin gene, *UBI4*, is derepressed only late in the growth cycle, during nitrogen starvation or after heat shock — all conditions under which ribosomal protein synthesis is repressed. Ubiquitin need not be derived from the fusion genes, however. If ubiquitin sequences are deleted from all the fusion genes, the cell still grows simply by turning on the polyubiquitin gene⁶.

The ubiquitin moiety of *UBI3* is helpful but not essential in the function of ribosomal protein S37. If S37 is expressed from a *UBI3* gene from which the ubiquitin-coding sequences have been deleted, many copies of the mutant gene are needed to replace fully one copy of the

UBI3 gene. One could speculate that the ubiquitin is used as a translational spacer for a very short protein, as a nuclear localization signal or as a 'wedge' to assist in assembling proteins into the ribosome. At what point in the assembly process is the ubiquitin cleaved? An alternative view derives from the observation that the ribosomal proteins turn over

very rapidly if not assembled into a ribosome^{11,12}. Ubiquitin could protect the ribosomal proteins from degradation, or perhaps cleavage from a ubiquitin molecule is a stratagem to make a protein without an amino-terminal methionine.

Although neither ubiquitin nor ribosomal protein seems essential for the other, there must be some explanation for this ancient and enduring marriage. That explanation may reveal the original role of ubiquitin, or some hitherto unsuspected connection between protein synthesis and protein degradation. □

Jonathan R. Warner is in the Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

1. Rechsteiner, M. A. *Rev. Cell Biol.* **3**, 1–30 (1987).
2. Wu, R.S., Kohn, R.W. & Bonner, W.M. *J. biol. Chem.* **256**, 5916–5920 (1980).
3. Ozkaynak, E. *et al.* *EMBO J.* **6**, 1429–1439 (1987).
4. Lund, P.K. *et al.* *J. biol. Chem.* **260**, 7609–7613 (1985).
5. Swindle, J. *et al.* *EMBO J.* **7**, 1121–1127 (1988).
6. Finley, D., Bartel, B. & Varshavsky, A. *Nature* **338**, 394–401 (1989).
7. Redman, K.L. & Rechsteiner, M. *Nature* **338**, 438–440 (1989).
8. Otaka, E., Higo, K. & Itoh, T. *Molec. gen. Genet.* **195**, 544–546 (1984).
9. Rotenberg, M.O. & Woolford, J.L. *Molec. cell. Biol.* **6**, 674–687 (1986).
10. Woudt, L.P. *et al.* *EMBO J.* **5**, 1037–1040 (1986).
11. Warner, J.R. *J. molec. Biol.* **115**, 315–333 (1977).
12. Maicas, E.F. *et al.* *Molec. cell. Biol.* **8**, 169–175 (1988).

Origins of marginal basins

*Leg 124 shipboard scientific party**

THE Sulu and Celebes seas are tectonically critical, but poorly known, marginal basins in the western Pacific Ocean. On Leg 124, completed in early January, we sought to determine the age, stratigraphy, palaeo-oceanography and current state of stress in the basaltic basement of these basins, drilling at five sites (see figure). We discovered unusually fine magnetostratigraphy in the basins, including the documentation of a magnetic reversal event within the Matuyama chron (approximately 1.1 million years (Myr) ago), which has been observed in few locations previously.

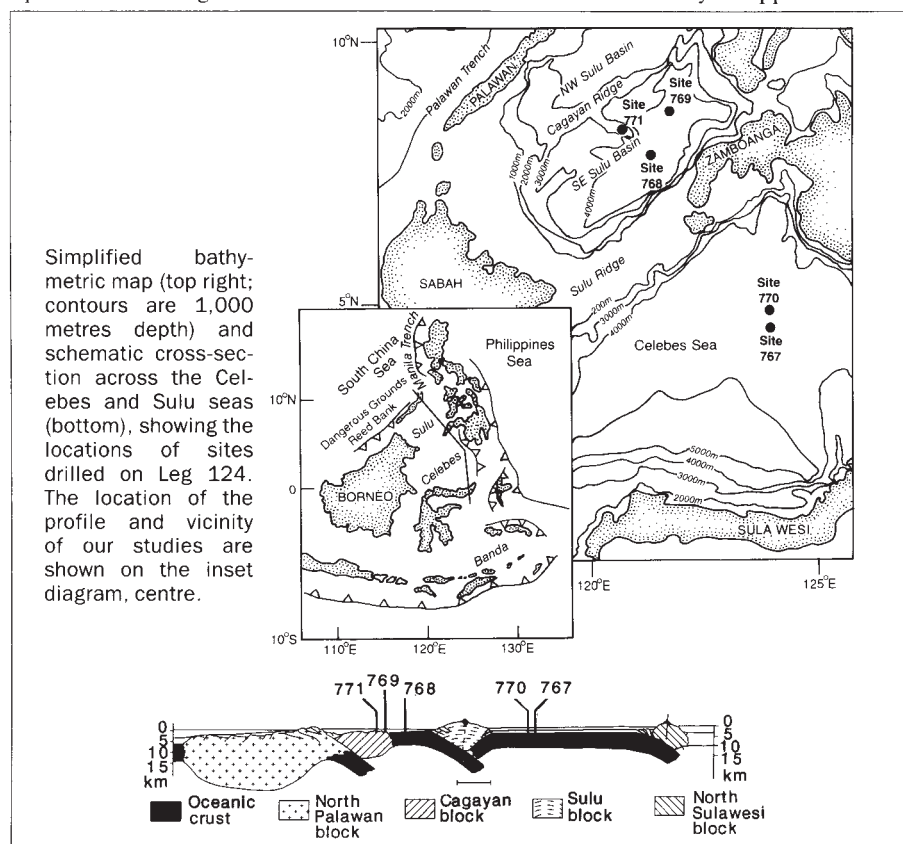
The western Pacific region is characterized by a series of marginal basins bordered by festoons of island arcs. The Sulu, Celebes and Banda basins are thought by some to represent an originally continuous, trapped fragment of a once-larger ocean basin, and by others to have formed by back-arc spreading. Our results rule out a single origin for all the basins, and it seems most likely that the origins of the Celebes and Sulu basins are very different. The basins lie in a broad zone of collisional tectonics, and one of our principal objectives was to use the sedimentary record of the basins to constrain the history of collisional events. Also, stress orientations within the basins establish the primary forces active in this complex zone of collisional tectonics, and the comparison of these results with those from Leg 123 in the south-east Indian Ocean should be instructive.

Basement underlying the Celebes Sea is plagioclase-olivine pyritic basalt, and the Sulu Sea basement is composed of olivine basalt. We recovered pillow basalts at both sites in the Celebes Sea — 100 m at site 770 but only 42 cm at site 767. The cores from site 770 include sill and glass-rimmed pillow basalts and intercalated massive, veined flows. We drilled to a depth of 222 m at site 768 in the Sulu Sea basement, recovering a series of pillow basalts and sheet flows, intruded by

two dolerite sills. The trace-element geochemistry of the Celebes Sea basement (site 767) bears the signature of normal mid-ocean ridge basalts (MORB), whereas the Sulu Sea basalts seem to be transitional between MORB and island-arc tholeiites.

The stratigraphies of sediments in the basins yield a detailed record of their histories. The Celebes Sea originated in an open ocean setting in the middle Eocene

as a back-arc or intra-arc basin in the late early to early middle Miocene as determined by radiolarian biostratigraphy in sediments overlying basement. This age coincides with the collision of the Cagayan Ridge with the rifted continental margin of China. Initial deposition of volcanic brown clay was followed shortly thereafter by coarse tuffs from rapidly deposited pyroclastic flows. Slow accumulation rates (9 m Myr⁻¹) marked the early part of the middle Miocene, but in the late middle Miocene (10.5 Myr ago), very rapidly deposited (300 m Myr⁻¹) continentally derived turbidites inundated the basin. These volcanic ash layers appeared in the



(42 Myr ago). This age is constrained by the biostratigraphy of clays bearing radiolarians (planktonic protozoa) directly above basement. This determination is consistent with previously determined magnetic anomalies in the basin (Weissel, J.K. *Am. Geophys. Un. Geophys. Monogr.* 23, 37–47; 1980). By the middle Miocene (14 Myr ago) the sea was receiving continentally derived turbidite deposits rich in quartz and plant debris. By the late Miocene (6 Myr ago) nearby explosive arc volcanism was producing thin ash layers in the basin, and continued to provide a major component to the basin sediments until recently. No major carbonate units occur at site 767, but thin carbonate turbidites were deposited sporadically in the middle to late Miocene and more commonly in the Pliocene (5.1–1.6 Myr ago) at this site.

The Sulu Sea seems to have originated

basin 6 Myr ago, indicating the initiation of some of the volcanoes that presumably surround the basin.

A significant change in sediment type from green clay below to nannofossil marl above occurred 1.9 Myr ago, probably indicating the time of rapid shallowing of the surrounding sills, isolating the basin from deep waters of the Pacific. This was also a period of rapid shallowing of the calcite compensation depth worldwide, which could also have contributed to some of the change in sedimentation in the Sulu Sea.

Sites 769 and 771 on the flank of the Cagayan volcanic ridge, penetrated a section similar to that at site 768 in the Sulu Sea, but devoid of turbidites, and therefore very much thinner (280 and 240 m of sediments, respectively). Brown claystone at the base of the section at site 769 contains late early to early middle

* Co-Chief Scientists: Eli Silver (Univ. California, Santa Cruz) and Claude Rangin (Univ. Pierre et Marie Curie, Paris). ODP Staff Scientist: Marta von Breyman (ODP, Texas A&M Univ.). Also: Ulrich Berner (Federal Institute of Geosciences and Natural Resources, Hannover), Philippe Bertrand (CNRS, Univ. Orleans), Christian Betzler (Geological Inst., Tübingen), Garrett W. Brass (Univ. Miami), Vindell Hsu (Louisiana State Univ.), Zehui Huang (Dalhousie Univ.), Richard Jarrard (Lamont-Doherty Geological Observatory), Stephen Lewis (US Geological Survey, Menlo Park), Braddock K. Linsley (Univ. New Mexico), Dean Merrill (Texas A&M Univ.), Carla M. Müller (Univ. Frankfurt), Alexandra Nederbragt (Free Univ., Amsterdam), Gary Nichols (University College, London), Manuel Pubellier (Univ. Pierre et Marie Curie, Paris), Fernando G. Sajona (Mines and Geosciences Bureau, Quezon City), Reed P. Scherer (Ohio State Univ.), Der-Duen Sheu (Univ. Oklahoma), Hidetoshi Shibuya (Univ. Osaka Prefecture, Sakai), Jih-Ping Shyu (Texas A&M Univ.), Randall Smith (Tulane Univ.), Terence Smith (Univ. Windsor), Renato U. Solidum (Philippine Inst. of Volcanology and Seismology, Quezon City), Piera Spadea (Univ. Udine) and Dwayne D. Tannant (Univ. Alberta).

Miocene radiolarians, and marl at site 771 contains radiolarians and nannofossils of this age, the same as that of the Sulu Sea and the cessation of volcanism on Cagayan Ridge. We also found volcanic tuffs at site 769 beneath the brown claystone, but these tuffs are basaltic to andesitic in composition, indicating a different source from the rhyolitic to dacitic tuff in the deep Sulu Sea.

Logging at sites 770 and 768 included use of the borehole televiwer, which measures the shape and roughness of the hole in the basement. The holes tend to be elliptical, with their long axes lying perpendicular to the direction of maximum horizontal stress. In places the long axes tend to develop small 'breakouts', which are easily visible in the record. The method thus gives a measure of stress direction. Preliminary results at both sites indicate that the maximum horizontal stress direction trends to the north-east, which is consistent with the expected stresses related to the Miocene and Pliocene collisions of the Sulu, Cagayan and Palawan ridges with the Philippine mobile belt.

PALAEONTOLOGY

Sea-dragons all aswim

Michael A. Taylor

THREE major finds of fossil marine reptile on display in a new exhibition* in Bristol will fuel the current revolution in marine reptile studies. The existence of several different kinds of large predator in the same sea poses difficult questions about their ecological partitioning. Two of the new finds are ichthyosaurs from the Lower Liassic (about 200 million years ago) and the third is the skull of a pliosaur, from the rather younger Kimmeridge Clay of Westbury, Wiltshire, UK (see figure).

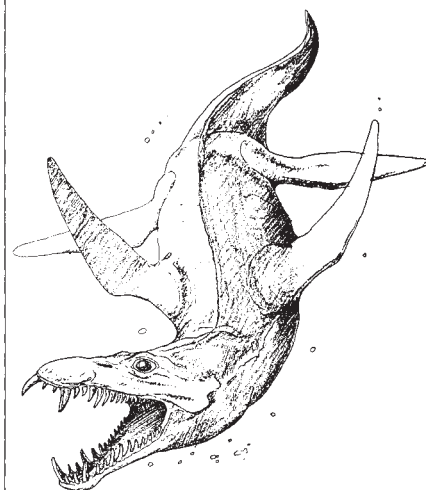
A virtually complete skeleton of an ichthyosaur from Charmouth¹ in Dorset, UK was about 9 m long in life. Although comparable in length with other large Liassic ichthyosaurs, its long, delicate snout and tiny, sharp teeth contrast with their robust snouts and heavier teeth. The large size of the Charmouth ichthyosaur combined with a narrow snout suggests a diet of small, abundant prey near the base of the food chain. But marine reptiles never seem to have exploited filter-feeding on even smaller prey, as do the modern mysticete whales, filter-feeding sharks and the newly reinterpreted giant teleost fish *Leedsichthys*² from the Jurassic Period.

The only known specimen of the much smaller ichthyosaur *Excalibosaurus costini*³ is also on display in Bristol. This

A short, normally polarized, full-reversal event within the Matuyama chron is evident in the palaeomagnetic record. Occurring just below the Jaramillo sub-chron 1.1 Myr ago, this reversal corresponds to the Cobb Mountain event documented by E.A. Mankinen *et al.* (*Geology* **6**, 653–656; 1978) and is supported by marine magnetic-anomaly profiles described by D.K. Rea and R.J. Blakely (*Nature* **255**, 126–128; 1975).

Maturity of the terrestrial type of organic matter at Site 768 reaches the stage of thermal hydrocarbon generation at a shallow sub-bottom depth, indicating a thermal gradient greater than 100 °C km⁻¹. We even observed some thermogenic generation of gas at this site.

Calcium and magnesium in interstitial waters show positively correlated increases with depth in the lower parts of the sedimentary sections of sites 767–770. Because normal alteration products of basalt scavenge magnesium, these observations suggest the presence of a chemical mechanism of crustal alteration previously unsuspected in the oceanic crust. □



Restoration of the pliosaur *Liopleurodon macromerus*, a 7.5-m-long oceanic predator of 150 million years ago (courtesy of John G. Martin and Bristol City Museum and Art Gallery).

of fast teleost fish has been implicated in the marine reptiles' decline: many had disappeared from the seas long before the end of the Cretaceous period (65 million years ago).

The exhibition celebrates Bristol Museum's acquisition and curation of these three major finds. But it also narrates the museum's key role in the history of research on marine reptiles, and laments the destruction, in an air raid in 1940, of the original collection of marine reptiles, the finest in Britain outside London^{12,13}. It was here during the 1820s that W. D. Conybeare and H. T. De la Beche carried out the first accurate assessments of ichthyosaurs and plesiosaurs, interpreting them as links in the "great chain of being" between fish and reptiles¹⁴. These marine reptiles helped to popularize palaeontology, and — ironically — to give the impetus leading to the replacement of the paradigm of the great chain by darwinian evolution. Now the cycle is complete and the marine reptiles are once again receiving well-deserved attention. □

Michael A. Taylor is at the Leicestershire Museum, 96 New Walk, Leicester LE1 6TD, UK.

animal, from Somerset, is distinguished by a long, protruding upper jaw, presumably for slashing through schools of prey too small to catch individually.

The Westbury pliosaur is a fortunate exception to the destruction of fossils by modern quarry machinery before they are even spotted⁴. The 1.6-m-long skull of *Liopleurodon macromerus* is exceptionally well preserved, and comes from an animal perhaps 6 to 7 m long. The robust jaws, strong pointed teeth with cutting keels, and wide gape indicate a predator which killed and tore apart any prey which it could catch^{5,6}.

The ecological diversity of ancient seas must have been very great to have supported all these large, carnivorous sea-dragons simultaneously. The next task is to sort out the adaptations of each in order to gain a better appreciation of their ecological niches. One approach is to use tooth form and wear, supplemented by stomach contents, to assess dietary specialization⁶. Another is to compare locomotor adaptations^{7–10}: the fish-shaped ichthyosaurs and the compact, short-necked plesiosaurs were probably fast pursuit predators, while the longer-necked plesiosaurs may have used their long necks to ambush smaller prey while cruising. Despite the presumed speed of ichthyosaurs and plesiosaurs, they were probably slow when compared with modern forms such as toothed whales^{9–11}. The evolution

*The Great Sea Dragons, City of Bristol Museum and Art Gallery, Queen's Road, Bristol BS8 1RL, 17 February–6 May 1989.

1. Clark, R.D. *The Charmouth Ichthyosaur. A Dorset Giant* (City of Bristol Museum and Art Gallery, 1989).
2. Martill, D. M. *Neues Jb. Geol. Paläont. Mh.*, 670–680 (1988).
3. McGowan, C. *Nature* **322**, 454–456 (1986).
4. Swansborough, S. A. *The Westbury Pliosaur. A Jurassic Jaws* (City of Bristol Museum and Art Gallery, 1989).
5. Taylor, M. A. *Zool. J. Linn. Soc.* **91**, 171–195 (1987).
6. Massare, J. M. *J. vert. Paleont.* **7**, 121–137 (1987).
7. Halstead, L. B. *J. geol. Soc.* **146**, 37–40 (1989).
8. Riess, J. *Palaeontographica* **A192**, 93–155 (1986).
9. Massare, J. A. *Paleobiology* **14**, 187–205 (1988).
10. Taylor, M. A. *Palaeontology* **30**, 531–535 (1987).
11. Reif, W.-E. *Neues Jb. Geol. Paläont. Mh.*, 361–379 (1988).
12. Clark, R.D. *Dragon Hunters* (City of Bristol Museum and Art Gallery, 1989).
13. Clark, R.D. *Dragon Keepers* (City of Bristol Museum and Art Gallery, 1989).
14. Taylor, M. A. & Torrens, H. S. *Proc. Dorset nat. Hist. archaeol. Soc.* **108**, 136–148 (1987).

Painting by resonance

Chris Scarre

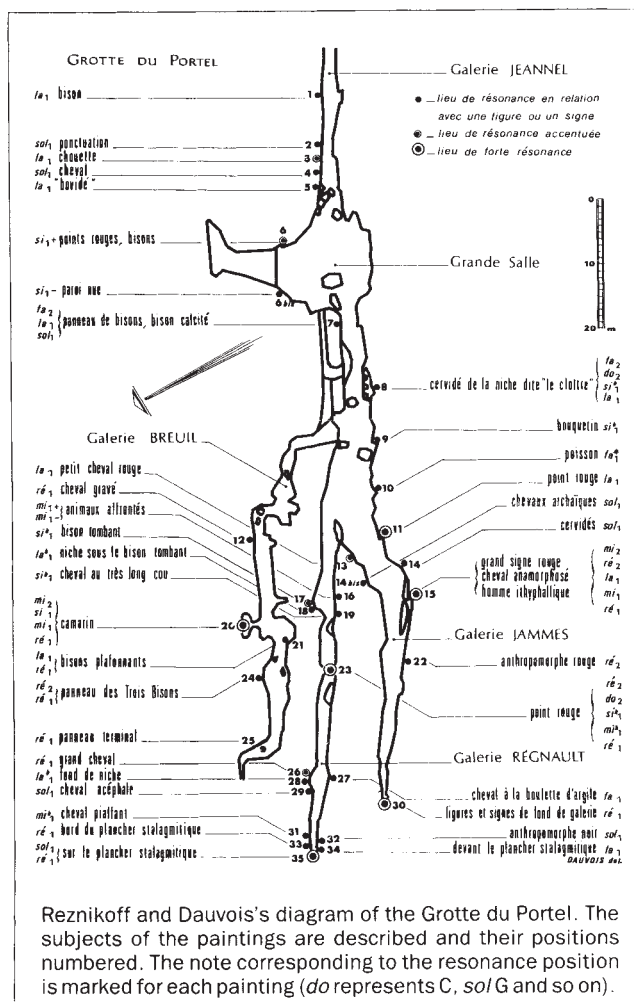
DID the painted caves of western Europe once resound to the music of Palaeolithic chants? Such is the thesis put forward by Iégor Reznikoff and Michel Dauvois in the latest issue of the *Bulletin de la Société Préhistorique Française* (85, 238–246; 1988). The authors have studied three caves in the Ariège department at the foot of the French Pyrenees. Their results suggest that the acoustics of the caves played a significant part in determining where the paintings were located, and this observation leads directly to the supposition that music or chants were important elements in cave ceremonies around 20,000 years ago.

Reznikoff and Dauvois rely on the fact that in certain places ("points of resonance") the caves resonate in response to particular notes. They proceeded slowly through the cave using their voices to produce a series of notes spanning almost three octaves, from C₁ to G₃. They extended the range of notes for a further two octaves by harmonics and whistling. Where there was a resonance response, they recorded the location and the particular note eliciting the response. They used these observations to draw up a resonance map of the cave (see figure).

The resonance of the caves is not in itself surprising, but the significance of the study becomes apparent when the authors compare their points of resonance with the location of cave paintings. They draw three main conclusions. First, most of the cave paintings are at or within 1 metre of points of resonance. The Grande Salle at Portel, for example, which gave no resonance response, also has relatively few paintings. Second, most of the points of resonance correspond to locations with cave paintings. Indeed, the best points of resonance are always marked in this way. Finally, the authors claim that the location of some of the paintings can be explained only by the resonance of that particular location. A good example is number 23 at Portel (see figure), where a particularly effective

point of resonance is marked by red painted dots, as there is not enough room for a full painted figure.

Reznikoff and Dauvois remark from their own experience on the impressive effect of cave resonance, which would have been all the more striking in the flickering half-light of the simple lamps or tapers used by the original artists. Drums, flutes and whistles may have been used in cave rituals — bone flutes have been found at several Palaeolithic sites in Europe of roughly the same age as the paintings. The potential of cave resonance



would, however, be elicited only by the much greater range of the human voice. The image of the cave artists chanting incantations in front of their paintings may not be too fanciful. Reconstructing prehistoric sounds is inevitably a risky and ambitious venture, but this study is of particular value in drawing new attention to the likely importance of music and singing in the rituals of our early ancestors. □

Chris Scarre can be contacted at the Department of Archaeology, University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK.

Time out of mind

LIKE many busy people, Daedalus often wishes there were more hours in the day. Indeed, he has given serious thought to slowing down the rotation of the Earth, but still cannot think of a suitable dumping-ground for the surplus angular momentum. Instead, he is exploring the possibility of speeding up the human internal clock. In most people its free-running period is about 25 hours, entrained with more or less reluctance to the 24-hour day by the pitiless light of morning. But those infuriating eager-beavers who snap into action at the crack of dawn presumably have a cycle of less than 24 hours, and so adjust to the natural day with some phase-lead.

Many other creatures also have an inbuilt circadian rhythm. It seems to be genetically determined; hamsters and fruitflies sometimes occur in mutated forms with a different circadian period, or even with none at all. These mutations seem to work by altering the output of some time-regulator protein. If the same mechanism works in man, then the human internal clock could be neatly adjusted by altering the output of this protein. So DREADCO's pharmacologists are trying to develop a drug which either blocks the receptors for the timing protein, or inhibits or facilitates its production. In the current state of knowledge, this has to be a hit-and-miss business; but the hormone melatonin and the benzodiazepine tranquilizers, both of which affect the biological clock, seem a good place to start. Serum analysis of volunteer sluggards and human dynamos may also be revealing.

DREADCO's Regulator® tablets will bring wonderful harmony to all our lives. Those who (like Daedalus) find that morning comes too soon, will shorten their cycle by cautious doses of the appropriate Regulator until they are in perfect tune with the clock. Their short-cycle brethren will lengthen their subjective day to make the most of the evening. But Daedalus is most intrigued by the possibility of a special Regulator which, like the *per⁰* mutation in fruitflies, does not merely alter the circadian rhythm but actually abolishes it. Acyclic individuals will live a very unusual life. They will show no particular tendency to go to sleep; once asleep, however, they will show no particular tendency to wake up. Having no time-sense, they will never be bored; and they will only eat when actually hungry. Acyclic workers will take over all the continuous industrial processes which are currently manned imperfectly, and often dangerously, in shifts. They will severely test theories about the importance of sleep and dreams to human mental functioning. And if our lifespan is indeed counted off in periods of the circadian rhythm, they should live forever.

David Jones

signals resulting in proliferation and differentiation of the B and T cells, and histamine release from the mast cells. A similar signal might be generated by the viral protein gp30 explaining the polyclonal B-cell proliferation (persistent lymphocytosis) connected with BLV infection.

MICHAEL RETH

Max-Planck-Institut für Immunbiologie,
D-7800 Freiburg,
FRG

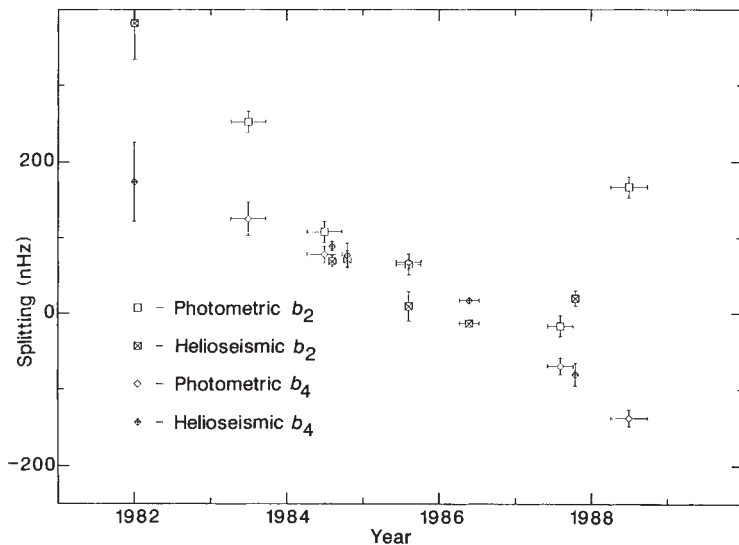
1. Clevers, H., Alarcon, B., Wileman, T. & Terhorst C. A. *Rev. Immun.* **6**, 629–662 (1988).
2. Weissman, A.M. et al. *Science* **239**, 1018–1021 (1988).
3. Sakaguchi, N., Kashiwamura, S., Kimoto, M., Thalman, P. & Melchers, F. *EMBO J.* **7**, 3457–3464 (1988).
4. Hombach, J., Leclercq, L., Radbruch, A., Rajewsky, K. & Reth, M. *EMBO J.* **7**, 3451–3456 (1988).
5. Blank, U. et al. *Nature* **337**, 187–189 (1989).
6. Burny, A. et al. *Cancer Surv.* **6**, 139–159 (1987).
7. Sagata, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **82**, 677–681 (1985).
8. Hermanson, G.G., Eisenberg, D., Kincade, P.W. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6890–6894 (1988).

Solar oscillation

SIR—Gough discusses the recent helioseismic evidence for solar-cycle variations in the global solar aspherical structure. But he does not mention that the variable, even, helioseismic-splitting coefficients; the solar-limb temperature measurements; and the solar-constant data together yield a consistent interpretation of solar asphericity. Gough speculates that the splitting coefficients are directly related to sunspot number, and using a Legendre polynomial decomposition of the surface distribution of sunspots, makes a prediction¹. It is not obvious, however, that the sunspot number (surface density) and frequency splittings should be plotted on the same graph. This approach contrasts with my earlier calculation^{2,4} which explicitly predicts the splittings from an asphericity in the sound speed. These model predictions have been confirmed by observation.⁵

In his report¹, Gough criticizes the limb temperature observations and inferences because they do not account for the variability of the mean solar irradiance measured by ACRIM. In fact the limb data nicely explain⁶ the observed cycle variation of the solar constant. Furthermore, the formalism and data directly provide information on the depth dependence of the asphericity. For example, I have used the helioseismic observations to show that the asphericity is best described by a perturbation in the outer superadiabatic layer of the convection zone³. Thus, unlike Gough, I find it hard to see how the helioseismic data can be interpreted to explain a measureable solar-cycle variation in the neutrino flux from the Sun's core.

In any case, Gough's 'prediction' may soon be tested. Splitting data from 1988 should show large changes (and significant centroid frequency shifts) from 1987. In contrast with Gough¹ I expect the b_4 (α_4 in ref. 2) splitting coefficient to be even



Time dependence of the helioseismic splitting measurements^{3,4}. The most recent photometric observations are from limb photometry obtained in 1988. In contrast with Gough², I expect the b_4 coefficient to decrease in 1988.

smaller than in 1987 (see figure). Note that this is not a correlative inference, but a test of a simple model consistent with observed solar irradiance variations.

JEFF KUHN

Department of Physics and Astronomy,
Michigan State University,
East Lansing, Michigan 48827, USA

GOUGH REPLIES — That the even splitting coefficients of solar 5-minute oscillations should arise predominantly from a latitudinal variation in the propagation speed c of acoustic waves rather than a geometrical distortion of level surfaces is a natural supposition. The frequency splitting $\delta\nu/\nu$ between sectoral and zonal modes was $\sim 2 \times 10^{-4}$ in 1982 near the solar maximum⁷, whereas the oblateness at that time was probably no greater than 2×10^{-5} (ref. 8). Moreover the frequency splitting induced by the oblateness is of opposite sign to that observed.

What is the nature of the variation in c ? The most natural thought is that it is a sound-speed variation, associated with a variation in temperature. Simple models of the solar envelope in which thermal asphericity is induced by magnetic inhibition of convection in the equatorial regions, yet which maintain radial hydrostatic balance, reproduce the splitting data⁹. But they require a photospheric temperature variation that is a hundred times greater than recent observations indicate⁶. What is remarkable, however, is Kuhn's finding^{3,4} that if the observed latitudinal variation in relative temperature⁶ away from the poles is assumed to extend essentially unmodified through the superadiabatic convective boundary layer, then the splitting data are reproduced.

It was the original failure of plausible hydrostatic-envelope models to explain the splitting data that induced an investigation of magnetic effects, particularly as the asphericity was similar to the sunspot

distribution. A non-uniformly distributed fibril magnetic field of magnitude similar to that observed reproduces the required frequency splitting⁷. It was that result that led me previously¹ to assume a correspondence between magnetic activity, measured by sunspot density, and $\delta c/c$.

From an extrapolation of the sunspot data from 1 January 1988 I thus predicted the results of measurements being carried out in Antarctica by a team from the US National Solar Observatory. Recently, R. K. Ulrich kindly provided me with Mt Wilson sunspot data up to 25 November 1988. Using the scaling from my letter¹ these data provide a revision of the prediction to $\alpha_4 \approx 170$ nHz, $\alpha_4 \approx -110$ nHz; thus in contrast with my previous extrapolated result α_4 seems still to be declining, in agreement with Kuhn's expectations.

The new discussion of limb brightness measurements⁶ was not available when I wrote my first report¹. The impressive correspondence with the ACRIM data lends considerable credence to the interpretation in terms of effective-temperature variations, and thus demands a consistent theoretical investigation of its implications concerning the dynamical balance of the convection zone and the splitting of oscillation frequencies.

D. O. GOUGH

Institute of Astronomy,
University of Cambridge,
Cambridge CB3 0HA, UK

1. Gough, D.O. *Nature* **336**, 616–619 (1988).
2. Gough, D.O. *Nature* **336**, 720 (1988).
3. Kuhn, J.R. in *Seismology of the Sun and Sun-like Stars* (ed. Rolfe, E.) (ESA, Noordwijk, 1989).
4. Kuhn, J.R. *Astrophys. J.* **331**, L131–L134 (1988).
5. Jefferies, S.M., Pomerantz, M., Duvall, T.L., Harvey, J.M. & Jaksha, D.B. in *Seismology of the Sun and Sun-like Stars* (ed. Rolfe, E.) (ESA, Noordwijk, 1989).
6. Kuhn, J.R., Libbrecht, K.G. & Dicke, R.H. *Science* **242**, 908–911 (1988).
7. Duvall, T.L. Jr et al. *Nature* **321**, 500–501 (1986).
8. Dicke, R.H., Kuhn, J.R. & Libbrecht, K.G. *Astrophys. J.* **318**, 451–458 (1987).
9. Gough, D.O. & Thompson, M.J. *Advances in Helio- and Asteroseismology* (eds Christensen-Dalsgaard, J. & Frandsen, S.) 175–180 (Reidel, Dordrecht, 1988).

A way with words

Stephen Jay Gould

The Oxford English Dictionary, 2nd edn. Prepared by J.A. Simpson and E.S.C. Weiner. Clarendon: 1989. Twenty volumes, £1,500, \$2,500.

THE Brothers Grimm, after a lifetime of service collecting folk tales and developing some basic principles of linguistic evolution, began their greatest project in 1838 — a German dictionary that would illustrate the history of all words found in three centuries of literature, from Luther to Goethe. As the architects of mediaeval cathedrals understood so well, some projects are too big to complete in a lifetime, and satisfaction must reside in a worthy start and a workable plan. Wilhelm, who died in 1859, only reached the letter D, while Jacob, who lived four years longer, got all the way to F. Their great work was finished in our own century.

When James A. H. Murray became the first editor of *The Oxford English Dictionary* in 1879, he faced an even more daunting task — for his project would document all English words in all meanings, active and extinct, from 1150 up to the moment, and including actual quotations for all nuances of use. (The starting point was arbitrary, but sensible as a criterion for the inception of English, allowing about a century after the conquest of 1066 for the Romance language of the victorious Normans to fuse sufficiently with the indigenous Germanic of the vanquished Anglo-Saxons.) Murray anticipated a labour of ten years, and a product of four volumes and 6,400 pages. In 1884, with half his time elapsed, Murray published the first volume, reaching halfway through the first letter to 'ant'. But, as Aesop's grasshopper discovered, industry is an ant's middle name. Murray and his worldwide network, ranging from full-time editors to volunteers from Madras to California, persisted. The task was finished in 1928, long after Murray's death.

A project such as the *OED* cannot end so long as the genealogical skein of English-speaking people persists — for language is the hallmark of humanity, and linguistic change is the primary record of history. Any stated completion could only be an admission of failure. At most, the guardians of this greatest monument in the history of English scholarship may choose, now and again, to mark a way-

station in their eternal updating by issuing a hard-copy version of their dynamic storehouse. (Such overt summations may become less and less necessary as computer technology permits continuous change in electronic versions, with immediate access to all revisions and additions as they occur; though a few die-hards who love the heft, feel and smell of a real volume may — God preserve them forever — keep the art of bookmaking alive.) This second edition, in 20 volumes of 21,728 pages, defining about half a million words with illustrations from about 2.4

million quotations, is both a progress report and one of the greatest causes for celebration in the history of publishing.

The *OED* is timeless in the ethical sense of possessing ultimate and permanent value *in se*. But it is also the child of its time in form and purpose. It began as a project of the Philological Society in 1857 (an effort that delivered nearly 20 years of preliminary work to Murray's later editorship — and he still took five years to reach ant!). Two years later, in 1859, the Society published guidelines for the great work, including two fundamental principles: that "it should contain every word occurring in the literature of the language"; and that "in the treatment of individual words the historical principle will be uniformly adopted". This crucial decision, the foundation of the *OED* ever since, reflects the most enduring and innovative theme of nineteenth-century scholarship in general — the use of historical connection and development as the chief principle of organization and explanation.

Geology had burst the bonds of time and provided an ample stage; biology had discovered the genealogical basis of all natural order; anthropology and linguistics were reconceived as historical sciences. It is no accident that these guidelines for the *OED* were published in the same year as Darwin's *Origin of Species*.

In the *OED*, all words are treated as historical entities with origins in etymologies and histories of changing usage. The earliest meanings are presented first, whatever their current status, followed by later developments in genealogical sequence. Since words evolve in the manner of organisms, the editors faced the same problem confronting darwinian taxonomists who had to translate the actual geometry of evolutionary branching into the linear format of a book. They adopted the biologist's convention of listing primary meanings ('species') in linear order (by arabic numerals in the *OED*), but gathering them by groups into higher taxa ('genera'), each representing a separate branch (roman numerals in the *OED*).

This format frees the *OED* from a dictionary's conventional role as the arbiter of 'correct' usage. Standard modern definitions have no pride of place or special designation; they take their proper order in the historical flow. But this format also makes the *OED* the greatest reference book ever written — for it turns each word into a story, and its totality becomes our history, if not our very definition as human. In a world of contingency, no essence exists beyond the labyrinthine and

ant (ænt). Forms: 1 (*W. Sax.*) æmete, -ette, -ytte, 3-4 amete (amote), amte, 4-6 ampte, 5-6 ante, 5-7 annt, 6- ant. Also 1 (*Anglian*) *ēmete, 3-4 emete (-atte), 4-6 emote, 6 emmette, -otte (-ont), amyte, emet, 6-7 emmot(t, 6- emmet. *Pl.* ants (1 æmetan, 2-4 ameten, 4 amptes). [*OE.* æmete, ēmete, cogn. w. *OHG.* âmeiza, *WGer.* *āmaitjō, f. ā- (see *Æ-* pref.) off, away + *maitan*, *ON.* meita, *OHG.* meizan 'to cut,' as if 'the cutter or biter off.' (*Graff.*) The *OE.* became in 12-13th c. āmete or ēmete in different dialects; āmete has by suppression of medial vowel and bringing together of two consonants become amte (ampte), ante (cf. *account* for *accompte*), ant; ēmete, retaining the medial vowel, is now EMMET, q.v. *Ant* is the leading literary form.]

1. a. A small social insect of the Hymenopterous order, celebrated for its industry; an emmet, a pismire. There are several genera and many species, exhibiting in their various habits and economy some of the most remarkable phenomena of the insect world. (For other quotations see EMMET.)

c 1000 *Sax. Leechd.* I. 87 Æmettan ægru genim. 1297 *R. Glouc.* 296 As pycke as ameten crepeþ in an amete hulle. 1340 *Ayenb.* 141 Alsuo ase þe litel amote. 1382 *Wyclif Prov.* xxx. 25 Amptis [1388 amtis] a feble puple, that greithen in rep time mete to them. 1430 *Lydg. Chron. Troy* I. i. He sawe by the earthe lowe Of Antes crepe passing greate plente. 1533 *Elyot Cast. Helth* III. xii. 66b, The lyttelle ant or emote helpeth up his felowe. 1547 *Boorde Brev. Health* clxi. 58 Amytes, or Pysmars, or Antes. 1585 *Lloyd Treas. Health* B viij, Pouder of Amptes, myxte with Oyle. 1578 *Mascall Planting* (1592) 58 For to destroy Emets or Antes, which be about a Tree. 1611 *Bible Prov.* vi. 6 Goe to the Ant [Wycl. ampte, amte, Coverd. Emmet], thou sluggard. 1633 *G. Herbert Ch. Mil. in Temple* 184 The smallest ant or atome knows thy power. 1642 *Sir T. Browne Relig. Med.* 30 The wisdom of Bees, Annts and Spiders. 1733 *Pope Ess. Man* III. 184 The Ant's republic, and the realm of Bees. 1838 *Penny Cycl.* X. 372 Formic Acid, or acid of ants. 1861 *Hulme Moquin-Tandon* II. IV. i. 213 When the Red Ant (*Formica Rufa*) crawls over a piece of litmus paper, it produces a red track.

b. In colloq. phr. to have ants in one's pants (orig. U.S.), to fidget constantly, esp. because of extreme agitation, excitement, nervousness, etc.; to be impatient or restless. Cf. ANTSY a.

[1939 *Kaufman & Hart Man who came to Dinner* I. ii. 62 'I'm in love.'... 'I'll pull you out of this Miss Stardust... I'll get the ants out of those moonlit pants.' 1040 'C.



unpredictable paths of historical development. The *OED* is the ultimate companion for those lonely years on a desert island in our standard *Gedanken* experiment. We consult it for a definition, and invariably emerge with an adventure. Half a million stories to last for as many thousand-and-one nights as any person can endure. (I'm not sure that I would trade my *OED* even for Scheherazade herself.)

Start anywhere. 'Chemistry' began with two extinct meanings as alchemy and paracelsian versus galenic medicine. Its modern definition dates from the seventeenth century, but the vernacular meaning of "instinctual and unanalyzable attraction or affinity between people" is even older — dating from a comment of Queen Elizabeth (the first one) in 1600, and not to the beaches of modern California. 'Physics', thanks to Aristotle, is an even older word, but it encompassed all nature (even including God and spirits according to Locke) until eighteenth-century usage began to split the organic and inorganic worlds. 'Biology' is a relative newcomer, dating only from the German Treveranus and the Frenchman Lamarck in 1802 (with the first use in English in 1813).

'Science' (from the Latin *scire*, to know) goes back to the beginning, but as a general word for systematized knowledge of any sort (Chaucer wrote about "a science of good works"). The modern usage follows a series of branching and progressive confinement. Science began to separate from art in the eighteenth century, but restriction to knowledge about the physical and natural world only occurred in the mid-nineteenth century. Keeping abreast of the latest developments, the final definition presents 'science park', with an American source of 1972 for first usage.

Related words teach us more. 'Scientism' was used in a purely descriptive sense as early as 1877, but the modern derogatory meaning dates from G.B. Shaw in 1921 — in *Back to Methuselah*, and with a capital S ("The iconography and hagiology of Scientism are as copious as they are mostly squalid"). The concept of 'scientist' may date from the Domesday Book, but the word only goes back to 1834. The *Quarterly Review* then noted "the want of any name by which we can designate the students of the knowledge of the material world collectively". "This difficulty was felt very oppressively" at a recent meeting

of the British Association for the Advancement of Science. The nameless attendees considered philosopher as "too wide and lofty" and "savans" (*sic*) as "rather assuming". "Some ingenious gentleman" then suggested "scientist" by analogy with artist, claiming "no scruple" for the odd termination of *ist* by noting



James A.H. Murray, word collector extraordinary.

its acceptance in "economist" and "atheist". "But this was not generally palatable", the *Quarterly* reports, and scientist might have died aborning, if William Whewell had not taken up the cause in 1840. And a good thing too — or we might have been saddled with "scientman", proposed once in 1636, but quickly exterminated.

The three-page entry on 'nature' is a fascinating treatise on English thought about essential qualities (the original definition of properties held by birth, or natively). Category I speaks of the "essential qualities of things"; II of vital or physical powers in humans, including two extinct meanings ("semen or menses" in a lineage extending from Chaucer to James Joyce, and "the female pudendum, especially that of a mare" extending from Caxton through to the eighteenth century); III of the power or impulse directing or controlling human action (as opposed to grace); IV (now getting close to modern scientific usage) as regulative powers that control the physical world, or

even (but not until the seventeenth century), "the material world with its objects and phenomena". This last (and the basis, I assume, for the name of this journal), is category IV, definition 13 in the *OED*. Nature, by the way, has been personified as female ever since Chaucer.

As a vast collective project the *OED* is far from perfect, and inevitably so. But our complaints can only represent lovers' quarrels. I caught some vestiges of Victorian origins — references to the tools and habits of "African savages", for example. I found serious errors and omissions in the few words that I have studied intensely ('evolution' is badly served because the compiler did not grasp the full extent of pre-darwinian biological usage for the embryological theory of preformation). Some extinct words receive too much play (the complex terminological apparatus for Weismann's incorrect theory of heredity), while others now in vogue have not come to the *OED*'s attention (the far less complex terminology for evolutionary change in the timing of development).

This work needs no medals, no citations, no words of praise. It stands by and for itself, as a gift to all intellectuals. The truly great, in supreme and justified confidence, merely present themselves — *ecce homo, ecce res*. Christopher Wren has no gravestone, only a simple plaque on the floor of his greatest building, St Paul's Cathedral, engraved with the words: *si monumentum requiris, circumspice* (if you are searching for a monument, look around). For the *OED*, we need only say: *si monumentum requiris, inspice* — look within. □

Stephen Jay Gould is a Professor in the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Big business

Marco Festa-Bianchet

Megaherbivores: The Influence of Very Large Body Size on Ecology. By R.N. Owen-Smith. Cambridge University Press: 1988. Pp. 369. £40, \$64.50.

MEGAHERBIVORES are terrestrial herbivores that weigh over 1,000 kg: elephants, rhinos, hippos and giraffes. Biologically they are unique in several respects. For example, as adults they are almost completely beyond the reach of non-human predators. They also efficiently use low-quality forage; and because they need a lot of it, they have a great effect on ecosystems, and on the kinds and numbers of plants and animals that share their environment.

In his book devoted to these huge creatures, Owen-Smith first considers

their ecology, physiology and behaviour. In most cases, he presents a test of whether they are just especially large in size or are different from what may be predicted by extrapolating allometric relationships of smaller herbivores. He discusses the limitations inherent in the quality of the available data and in the statistics used to analyse them. Megaherbivores are difficult to study. Their longevity and slow reproductive rate, coupled with human exploitation and habitat destruction, cast uncertainty over the causes of particular ecological processes; for example, habitat fragmentation may prevent animals from dispersing and lead to severe habitat degradation. Perhaps in the past megaherbivores avoided this problem by moving to new areas. Exploitation by human beings probably affected population dynamics and habitat characteristics, so that it is sometimes unclear which changes in vegetation structure are a normal feature

of the community ecology of megaherbivores, and which result from the cessation of exploitation and sudden population increases in habitats whose structure changed while numbers were kept low.

The middle chapters of the book vary in scope. In some, for example those on nutrition, feeding ecology, reproduction and ecosystem processes, Owen-Smith can consider all megaherbivores as a group, and tests well-defined hypotheses. In others, such as those on behaviour and demography, differences between the species force less interesting species-by-species summaries of data. All chapters provide good literature reviews.

Most of the book is a prelude to two arguments presented at the end, where Owen-Smith puts forward explanations of the Pleistocene extinctions and discusses the conservation of surviving species. Megaherbivores, once distributed over all continents, are now restricted to eight species in Africa and tropical Asia. Only hippos and giraffes appear to be safe over most of their range. Yet local problems of overpopulation show that "megaherbivores are embarrassingly successful when protected from human predation".

The explanation for Pleistocene extinctions has human overexploitation as the main driving force, but takes into account climatic shifts and changes in habitat characteristics. Owen-Smith suggests that it was hunting that caused the demise of megaherbivores, especially in North America and Europe, and that changes in habitat (leading to extinction of some smaller species) were an effect and not a cause of their disappearance.

The chapter on conservation is excellent. Owen-Smith believes management is necessary for conservation, because artificial habitat fragmentation prevents dispersal of animals. He also recognizes that 'overgrazing' and 'overpopulation' are not biological realities, but reflect management choices. Protected from poaching, megaherbivores could recover to levels where light exploitation would be both feasible and desirable — but "without effective controls of illicit markets, populations of these great beasts will be driven inexorably towards extinction". □

Marco Festa-Bianchet is NATO Science Fellow in the Large Animal Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

New in Britain

• *What Mad Pursuit: A Personal View of Scientific Discovery* by Francis Crick. Publisher is Weidenfeld & Nicolson, price is £12.95. For review see *Nature* 336, 268 (1988).

New in paperback

• *Bones of Contention* by Roger Lewin. Publisher is Penguin, price is £5.99. For review see *Nature* 330, 277 (1987).



Defensive stare — detail from a bronze shield mount of the 2nd or 1st century BC found in the River Thames at Wandsworth. The picture is taken from *Celtic Art* by Ruth and Vincent Megaw, an illustrated survey of Celtic arts and crafts from 700 BC to AD 700. Published by Thames and Hudson, the book costs £28, \$45.

On course for a better life?

F.E.G. Cox

The Biology of Parasitism. Edited by Paul T. Englund and Alan Sher. *Alan R. Liss*: 1989. Pp. 544. Hbk \$90, pbk \$45. Distributed in Britain by Wiley, hbk £68.85, pbk £35.50.

EACH summer since 1980, a small group of young scientists has gathered at Woods Hole to attend a course on advances in the molecular biology and immunology of parasites. The course is given by acknowledged experts from all over the world, and serves as a barometer of the state of the subject. In this volume, a number of the lecturers have attempted to convey its flavour and spirit to a wider audience. In all, 41 scientists have contributed 28 chapters, grouped into biological, immunological and molecular biological, biochemical and genetic aspects of parasitism, to produce a book that attempts to reflect the present state of our knowledge and to be an essential aid in undergraduate and postgraduate teaching.

The stated aim is to emphasize concepts rather than review research data, but sadly some of the authors have ignored this principle and the volume is thus very varied in content. Things get off to an excellent start with accounts of the global impact of parasitic diseases by Ken Warren, and of parasitic zoonoses by George Nelson, which set the scene. There are also specialized chapters on human schistosomiasis, amoebiasis, American leishmaniasis and trypanosomiasis and, towards the end, on toxoplasmosis, that are stimulating and make

provocative points.

The first SDS-PAGE gel appears on page 97, closely followed by the inevitable contributions on the surface antigens of *Plasmodium falciparum*, thick with facts and thin on concepts. The surface antigens of trypanosomes also receive considerable attention in three chapters; in one of them, John Donelson manages to concentrate on essentials and to ask interesting questions, as does Alan Sher in his contribution on vaccination. Fresh ground is explored by C.C. Wang, who considers new targets for chemotherapy, but his is the only contribution on this important area. Most of the remaining chapters are largely mini-reviews of topics covered in depth elsewhere.

Overall, this collection of articles represents a fair overview of present-day parasitological research and pin-points the unevenness of the subject's coverage. The important matter is the extent to which knowledge gleaned in the main laboratories of the world can be applied to the immense problems outlined by Ken Warren, and it is disturbing that most of the questions asked are of an academic rather than a practical sort. Equally disturbing is that scrutiny of the faces in the photographs of those who attended the last eight courses reveals that only three are black. One is tempted to ask whether parasites are important because they are the playthings of immunologists and molecular biologists or because they are daily life-threatening components of the environment for millions of people. Whatever the answer, the editors have produced a book that is useful, readable and well worth having. □

F.E.G. Cox is a Professor in the Division of Biomolecular Sciences, King's College London, 26 Drury Lane, London WC2B 5RL, UK.

Measuring up the planet

Barry Parsons

Geophysical Geodesy: The Slow Deformations of the Earth. By Kurt Lambeck. Clarendon: 1988. Pp.718. Hbk £75, \$95; pbk £30.

GEODESY, which can claim to be the oldest of the geophysical sciences, maintains a somewhat different relationship to geophysics than the other geophysical sciences. The choice of *Geophysical Geodesy* as a title to emphasize that this book is about the geophysical applications of geodesy points to the usually dominant role of applied aspects of the subject. In contrast the title of a book on seismology would probably only be qualified if it were about exploration or engineering seismology.

This difference can be explained in part by the long timescales involved in the development of many geodetic techniques. Consider the example of very-long-baseline interferometry (VLBI), which is the method of determining the distance between two points on the Earth's surface by using radio telescopes at the two sites to measure the small differences in arrival time of signals from distant radio sources. It has taken 20 years of painstaking development for VLBI to get to the stage where it is possible to measure directly the relative motion between the tectonic plates. Although most geophysicists appreciate the potential of many geodetic measurements, their interest is only really aroused when it becomes clear that worthwhile information about the Earth will be forthcoming.

Fortunately for those interested in the geophysical applications of geodesy, in recent years several geodetic techniques have emerged that have either already demonstrated their value in studying the Earth or seem about to do so. Satellite altimeter measurements of the shape of the ocean surface have replaced direct measurements of gravity as the primary source of information about the gravity field over the oceans. These observations provide one of the few ways of looking at the effects of mantle convection beneath the oceanic lithosphere. In the next few years the Global Positioning System will become fully operational, and this method of measuring baseline lengths, which is similar to VLBI except that satellite radio sources rather than extraterrestrial ones are used, has the potential to determine exactly how crustal deformation in the tectonically active regions of continents takes place. The nature of the core-mantle boundary is the subject of much debate, and VLBI observations of the nutations of

the Earth's rotation axis allow estimates to be made of the departure of the ellipticity of the core-mantle boundary from the ellipticity expected due to the Earth's rotation alone.

Against this background, Kurt Lambeck has set himself the task of writing an account of the subject that encompasses not only the geodetic techniques but the geophysical applications as well. The breadth of material he covers is impressive, and the book will provide an excellent overview of the field for both students and research workers. Such a comprehensive treatment within a single volume has its drawbacks, however. The author has to assume that the intended audience of graduate students will already have been exposed to the basic ideas of both geodesy and geophysics, an assumption that will only be true for the more advanced

student. In addition, by introducing many results without the concomitant theoretical development, one loses a sense of the physics that lies behind the equations.

The subtitle, *The Slow Deformations of the Earth*, reflects the author's interests and the contents of the book are clearly weighted towards solid-Earth geophysics. Physical oceanography, which will be the principal beneficiary of the two satellite altimeter missions due over the next five years, appears in a subordinate role. Yet, with the broad range of geophysical applications that are included, and its appearance just as new observations are about to flood in, this book is bound to attract a wide audience. □

Barry Parsons is Reader in Geodesy in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

Today's data

Gunnar von Heijne

Computational Molecular Biology: Sources and Methods for Sequence Analysis. Edited by Arthur M. Lesk. Oxford University Press: 1988. Pp.254. £25, \$49.95.

FIVE years ago, in the typical molecular biology laboratory, there would be at least one computer whiz kid: a graduate student who stole time away from the bench to break into the central campus computer and entertain himself with adventure games. This student was often quickly assigned the more serious job of taking care of the sequence-analysis needs of the lab, installing the newly purchased software that no one else had found the time to get running. Thus everybody was happy: the student was promoted to the status of computer expert, and the 'real' scientists didn't have to soil their hands at the PC keyboard.

Times have changed. Even the staunchest 'wet' experimentalist has had to adapt to the fact that these days one simply cannot do molecular biology without occasional recourse to the help of a computer. This, of course, is mainly a result of the extraordinary growth of DNA sequence data over the past ten years: more than 20 million bases worth of sequence is now stored in the main data banks such as GenBank or the European Molecular Biology Laboratory's data library, and only computers can deal with such massive amounts of information.

The world of 'machine molecular biology' is not immediately accessible to the average biologist, however, and good introductory texts are badly needed. The so-called CODATA Task Group on Coordination of Protein Sequence Data

Banks has now put out a volume called *Computational Molecular Biology*, edited by Arthur Lesk, to meet the demand.

The book is organized around four questions — What data are available? How can one gain access to them and to the necessary programs? What calculations can be done? And how can the results of the calculations be intelligently and cautiously interpreted? The first and second questions are treated at length; the third and fourth are less fully explored. We thus learn a lot about a number of DNA- and protein-sequence data banks around the globe, and we are introduced to the hardware and software of sequence analysis as well as to some aspects of computer networking. But we are told less about the actual methods and algorithms that allow the computer to produce the output we want.

Lesk apparently sensed this bias at an early stage, and has written a number of the less 'computerish' chapters himself: chapters for people who care more about how to find probable coding regions, or predict the structure of a protein, than about the details of how a GenBank entry is organized. Lesk has also come up with a couple of remarks that I will cherish for a long time to come. Thus, in summing up secondary structure prediction for proteins, he delivers the following: "Running a secondary structure prediction on a newly-determined sequence just because everyone else does so, is to be deplored, and the fact that the results of such predictions are generally ignored is insufficient justification for doing and publishing them". This is the sort of text that should be printed on every sequence-analysis software diskette as a 'Warning from the Surgeon General'. □

Gunnar von Heijne is in the Department of Molecular Biology, Karolinska Institute Center for Biotechnology, Huddinge University Hospital, S-141 86 Huddinge, Sweden.

The search for the causes of breast and colon cancer

Walter Willett

Epidemiological studies of breast and colon cancers implicate diet as a causative factor but the evidence is stronger for colon cancer, the occurrence of which may be reduced by diets with less animal fat and more fruit and vegetables.

AMONG non-smokers, cancers of the breast and colon are the most important malignancies in Western societies. By the age of 75 years, over eight per cent of women in the United States will develop breast cancer and about three per cent of both sexes will be diagnosed as having colon cancer¹. Another one to two per cent of men and women will develop rectal cancer by the time they are 75; although this is sometimes not distinguished from colon cancer, it will not be considered here because some of its epidemiological features are distinct. Treatment of breast and colon cancer remains disappointing. Hence, considerable interest has focused on determining the causes of these cancers, with the hope that such knowledge will lead to practical means of prevention.

The occurrence of both breast and colon cancer is strongly associated with modern affluence, suggesting that they may share important aetiological factors. Indirect evidence, largely based on differences in cancer rates between countries and changes in rates among migrants, has implicated diet as potentially important, but specific aspects of diet have not yet been definitively identified from epidemiological investigations. Much of the available information is derived from case-control studies in which reports from cancer patients about their past dietary practices are compared with those of persons without cancer. The possibility of biased reporting of former diets in these studies is difficult to eliminate. Prospective cohort studies, in which dietary variables measured among a large number of people are related to subsequent risks of cancer, should provide more dependable data, but these are only beginning to be available because it takes time to accumulate enough cancer cases. The existing limited data suggest that, despite gross similarities in the epidemiology of breast and colon cancer, incidence of the colon cancer is likely to be more responsive to dietary change and thus be easier to prevent.

Roles of environment and genetics

Evidence is strong for a substantial genetic contribution to the risk of breast and colon cancers. Having a first-degree relative with breast cancer increases risk by about 1.5–3 fold^{2,3}. Although this familial association could conceivably be due to a shared environment, the much higher risks (approximately tenfold) associated with some familial patterns, such as having a mother or sister with bilateral breast cancer at a young age, makes it more likely that the familial associations are genetically based. Even one case of colon cancer in a first-degree relative increases the risk for other family members by approximately threefold^{4,5}.

Nevertheless, genetic factors are not the only explanation. Rates of breast and colon cancer vary strikingly between countries, by five to tenfold⁶, and migrants moving from low- to high-risk countries adopt the rates of the new country^{7,8}. Furthermore, there have been large increases in rates in some countries during this century^{9,10}. Thus, both breast and colon cancer almost certainly involve a strong interaction of genetics and environment. Without the necessary environment, few will develop these diseases; but in a suitable environment, those with a genetic predisposition will be at high risk. Aetiological

studies of breast and colon cancer will become far more powerful when specific genetic markers for predisposition become available.

Differences in rates between countries have been used to estimate that 90% of colon and 80% of breast cancers in the United States of America can be attributed to non-genetic factors¹¹. The challenge is to identify the specific environmental factors. Although there is a plethora of hypotheses, well established risk factors are few and do not account for the major international differences or lead directly to practical preventive measures. Diet is a prime candidate: Doll and Peto¹¹ have suggested that as much as 50% of breast and 90% of colon cancer in the United States might be prevented by changes in diet. Before examining specific dietary hypotheses, the better established risk factors will be briefly discussed.

Risk factors for breast cancer

As early as 1700, Ramazzini¹² observed that child-bearing reduces the risk of breast cancer. In an international study, MacMahon *et al.*¹³ found that the dominant protective factor was an earlier age at first birth, rather than the number of children: a first birth after the age of 35 years increased the risk about threefold compared with a first birth before the age of 20. Other data suggest that having a larger number of children also independently reduces the risk^{14,15}. The protective effect of early first full-term pregnancy may be due to a permanent differentiation of mammary stem cells, resulting in a reduced lifetime risk of breast cancer^{16,17}.

Many studies show an association between early age at menarche and increased risk of breast cancer; typically early menarche confers about a 1.5–2 fold increase in risk compared with late menarche^{14,18,19}. Early menopause provides protection^{14,20}; removal of both ovaries before the age of 35 reduces the risk of disease by about 0.6 compared with natural menopause. These associations with reproductive factors provide support for a hormonal role in the aetiology of the disease^{21,22}.

The relationship between sex hormone levels in blood or urine and risk of breast cancer have been examined in many studies but clear and consistent findings have not emerged. All these studies have been small, however, and most have been conducted among patients who already have cancer, who may have had altered hormone levels because of the disease. Key and Pike²³ concluded in a recent review that the data overall suggest a promoting effect of oestrogens and possibly also of progestagens.

The possible role of sex hormones in the development of breast cancer has raised concerns that oral contraceptives and post-menopausal replacement hormones might increase the incidence of this disease. Overall, little relationship has been seen between oral contraceptive use and breast cancer²⁴, but longer-term use before a first full-term pregnancy may elevate risk^{24,25}; more data are needed to resolve this important issue. In most studies, oestrogen-replacement therapy after menopause has not been associated with the incidence of breast cancer²⁶, but use of hormones for several decades may cause a modest

increase²⁷. Most investigations have concentrated on past users and the possibility that current use may promote tumour growth has generally not been addressed.

Ionizing radiation, including that of X-rays, is a well-documented cause of breast cancer, especially for exposure during adolescence^{28,29}, but this cannot account for a large proportion of the disease because not many women receive a significant exposure. Benign lesions are probably best considered as pre-malignant responses to primary aetiological factors: the subsequent risk of breast cancer is strongly related to the degree of proliferative change and atypia found in biopsies of such lesions³⁰.

Risk factors for colon cancer

Few specific risk factors for colon cancer have been established. Inflammatory bowel diseases, such as ulcerative colitis and Crohn's disease, greatly increase the risk³¹ but because of their rarity, these account for a very small proportion of cases. The rare familial polyposis syndromes are associated with high risk³¹, but also contribute only slightly to the overall incidence of colon cancer.

The presence of adenomatous colon polyps, which can be considered as precursor lesions³², is associated with an elevated risk of colon cancer. Because the polyps can be examined directly by endoscopy, they provide a way of studying the progression to colon cancer that does not exist for breast cancer.

Diet and cancers of the breast and colon

Dietary fat and breast cancer. For the last decade, the dominant aetiological hypothesis for breast cancer has been that high fat intake, and animal fat in particular, is the primary cause of the large differences in rates between countries³³. This notion is based largely on two lines of evidence: striking international correlations (approximately 0.80) between *per capita* consumption of fat and breast cancer rates⁸; and animal experiments in which diets high in fat increase the occurrence of mammary tumours. Subsequent evidence has not, overall, supported this hypothesis.

It is tempting to draw causal inferences from the international correlations, but these studies provide only weak evidence because the actual causes might involve many other factors related to economic development. Factors such as physical activity, reproductive variables, body composition at different times in life and energy balance are particularly difficult to measure and control in correlational analyses. A recent survey of cancer rates in 65 rural counties in China with similar low levels of economic development³⁴ strongly suggests that the international comparisons are seriously confounded by factors related to industrialization: fat intake ranged from 5–47% of energy but breast cancer rates did not vary with fat intake and were at most about one-tenth of those in the United States.

Case-control studies and prospective cohort studies provide more opportunity to measure and adjust for other factors that may be related to risk of cancer. For breast cancer, no material association with fat intake has been found in case-control studies from New York³⁵, Hawaii³⁶, Australia³⁷, Greece³⁸ and Japan³⁹. Although a study from Canada was originally reported as showing a positive association⁴⁰, it was not statistically significant and examination of the data indicates that the average fat intake reported by cases was virtually identical to that reported by controls. In other case-control studies, only a limited list of foods was included^{41–46}; some have construed sporadic associations with foods containing fat as evidence that fat causes breast cancer. These data are difficult to interpret because of the tendency to focus on foods within a study for which associations are seen and because other aspects of the diet, including total energy intake, cannot be controlled for.

The potential for distortion of associations because of differential recall of past diet is inherent in case-control studies but is eliminated in prospective investigations. Only two pros-

pective studies with a comprehensive assessment of diet have been published. In the larger, based on a dietary questionnaire completed by 89,538 registered nurses in the United States in 1980⁴⁷, weak inverse associations with incidence of breast cancer were seen for total, saturated, and polyunsaturated fat. In the other, there was a statistically significant protective effect of total dietary fat but this inverse relationship was weaker and no longer significant after controlling for total energy intake⁴⁸. In prospective studies with limited dietary data, a positive association was seen with meat intake in Japan⁴⁹ but not among Seventh-Day Adventists in California⁵⁰. In the latter, breast cancer rates tended to increase, rather than decrease, the longer a Seventh-Day Adventist vegetarian diet, which contains substantially less animal fat than the general diet of North Americans⁵¹, was consumed.

Findings from case-control and prospective cohort studies, which generally fail to support the dietary fat hypothesis, have been criticized on the basis that diets in the populations studied are not sufficiently heterogeneous and the methods for measuring diet are not accurate enough to detect potential associations^{52,53}. But even when the degree of measurement error associated with the dietary questionnaire used in the larger of the two prospective studies is accounted for⁴⁷, the data are compatible with only a very weak positive association with fat intake^{54,55}. Viewed another way, if the positive associations suggested by the international data were true, the inverse associations actually seen in the two prospective studies would be statistically extremely unlikely.

The published case-control and cohort studies cannot exclude the possibilities that reducing fat intake to well below 30% of total energy intake, the level now frequently recommended^{33,56}, or at an earlier period in life, might influence breast-cancer rates. No appreciable reductions in breast-cancer risk among vegetarian nuns⁵⁷ and Seventh-Day Adventists⁵⁸ have, however, been observed, which argues against an important influence of animal fat over several decades.

The relationship between fat intake and the occurrence of mammary tumours in animals, which cannot be reviewed in detail here, remains controversial^{59,60}. Restricting energy intake dramatically lowers the incidence of mammary tumours. As fat is uniquely dense in energy, a low-fat diet also tends to be low in energy unless the available energy intake is carefully controlled. Thus, rats on a low-fat *ad libitum* diet have lower tumour incidence than those on a high-fat *ad libitum* diet; in one typical study⁶¹, a reduction of 40% was seen. In the same study, however, when the high-fat diet was restricted to provide 20% less total energy intake, tumour incidence decreased by 90%.

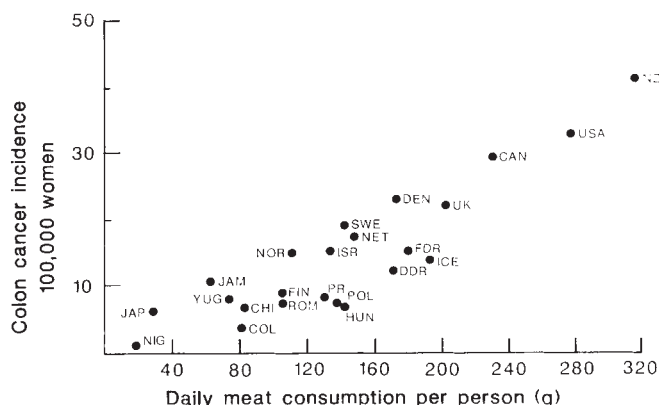


FIG. 1 Correlation between intake of meat per person and the incidence of colon cancer among women in 23 countries⁶. The strong positive association suggests that animal fat or protein may increase the risk of colon cancer. Because the countries with high meat intake and colon cancer rates tend to be affluent Western nations that differ in many ways from the countries with low meat intake and colon cancer rates, these data are far from conclusive.

Albanes⁶² has performed a meta-analysis of the experiments on diet and mammary cancer in mice over the past 50 years, in which he finds an extremely strong overall positive association for total energy intake. After adjustment for total energy intake, the fat composition showed a weak inverse relationship to incidence of mammary tumours. In another review of animal experiments, Birt⁵⁹ concluded that there is evidence for an effect of dietary fat which is independent of total energy intake.

Is any particular rodent model relevant to human breast cancer? The models usually involve a potent carcinogen or exposure to a mammary tumour virus, extreme contrasts in diet, and observations for relatively short periods before menopause. In a recent study⁶³ that more closely simulated the human experience, the occurrence of spontaneous mammary adenocarcinomas was investigated among more than 5,000 female rats and mice randomly assigned to either a 10% fat diet or corn-oil supplements that increased fat to 30% of calories. Even though the animals fed corn oil gained more weight than those on low-fat diets, the two-year cumulative incidence among rats was 2.5% on low-fat and 1.5% on high-fat diets. Among mice, the corresponding incidence rates were 1.7% and 1.3%. A case-control study of breast cancer in dogs, which was conducted by interviewing owners about the usual foods consumed by their animals⁶⁴, revealed an extremely wide variation in dietary fat composition but no association with risk of breast cancer. A consensus clearly does not exist at present that the total fat composition of the diet in animals influences the risk of mammary tumours, so laboratory data cannot be used as an argument for or against any relationship in humans.

The plausibility of the fat-and-breast-cancer hypothesis would be greatly enhanced if it could be shown that fat influences a biochemical mediator of breast cancer, but unfortunately no such mediator has been well established. Blood or urinary oestrogen levels are the most promising candidates, but the relationships between diet and hormone levels are unclear at present⁶⁵⁻⁶⁷.

Dietary fat and colon cancer. In contrast to breast cancer, epidemiological studies provide some, but not conclusive, evidence that animal fat or meat intake is associated with risk of colon cancer. In comparisons between countries, rates of colon cancer are strongly associated with intake of animal fat and meat (Fig. 1). High total fat intake has been associated with increased risk of colon or rectal cancer in case-control studies from Canada⁶⁸, Australia⁶⁹, Utah⁷⁰, England⁷¹, New York State⁷² and eastern Australia⁷³, although the last study is suspect as data from controls were not collected concurrently. In similar studies from southern France⁷⁴, Paris⁷⁵ and Belgium⁷⁶, however, no association was found. In the French and Belgium studies, intake of vegetable fats was inversely related to risk of colon cancer. The numbers of colon cancers in prospective studies have been relatively small. Among Japanese-Hawaiian men, those with the highest fat intake had a reduced risk of colon cancer⁷⁷ and among Chicago men no association was found between fat intake and colon cancer⁷⁸. In preliminary reports, animal fat intake was positively associated with risk of colon or large bowel cancer among the 89,538 women in the Nurses' Health Study⁷⁹ and among 35,000 Seventh-Day Adventists⁸⁰.

In case-control studies of diet and colon cancer, there is a notably consistent positive association with total energy intake⁶⁸⁻⁷². This is particularly puzzling because higher energy expenditure seems to reduce risk of colon cancer^{81,82} and obesity is unrelated, or at most weakly associated, with this disease (see below). The positive association with total energy intake has important implications for interpreting data on specific nutrients, because all tend to be correlated with total energy intake⁸³. It is important to adjust statistically for total energy intake in the analysis of these studies to distinguish between the effect of overall food intake and that of dietary composition, especially for nutrients such as dietary fat that are highly correlated with energy intake. Individuals must alter their intake of

specific nutrients primarily by manipulating the composition rather than the total amount of food eaten, unless they are to alter substantially their physical activity or weight. Thus, the intake of nutrients adjusted for total energy intake is most informative for making decisions about changes in diet. The association of dietary saturated fat with colon cancer was independent of total energy intake in one study⁸⁴ but not in another⁶⁹; most other case-control studies have not evaluated whether the observed association with overall fat intake is independent of energy intake.

Plausible mechanisms exist for an influence of dietary fat on the incidence of colon cancer; higher fat intake increases excretion of bile acids and the growth of colonic bacteria capable of converting bile acids to carcinogenic substances^{85,86}.

Dietary fibre

Burkitt's hypothesis⁸⁷ that dietary fibre (non-digestible complex carbohydrates) reduces the incidence of colon cancer seems to be, at best, an over-simplification. Increased intake of fibre is thought to reduce the transit time for faecal matter in the gut, and to dilute potential carcinogens in the faeces⁸⁸. High intake of fruits and vegetables has been consistently related to lower risk of colon cancer but consumption of cereal products has not. Although countries where there is a lower intake of fibre tend to have higher rates of colon cancer⁸⁹, inverse associations have been seen in some^{70,72,73,76} but not all^{68,74,69,71,75} case-control studies.

Some of the inconsistency may result from inadequate adjustment for total energy intake, for even the direction of association can depend on whether energy intake is appropriately considered in the data analysis⁷⁰. Evidence that grain products are protective is considerably weaker than it is for fruits and vegetables; in all eight studies in which the sources of fibre were examined separately, grain fibre or cereal intake was either unrelated or positively associated with risk of colon cancer, whereas intake of fibre from fruits and vegetables was protective^{69,74,82,90-94}. A protective effect of fruits or vegetables was also found in two additional case-control investigations^{95,96}.

Although the evidence that fruits and vegetables are protective is strong, the specific foods related to reduced risk are poorly defined and the responsible factors are unknown. Specific fractions of fibre may be the active agents; other candidates include vitamins, such as C or E, protease inhibitors⁹⁷ and indoles⁹⁸. In one analysis, no association was found with β -carotene⁹⁹, whereas an inverse relationship was seen in another⁹².

More limited evidence indicates that some vegetable factor may be protective for breast cancer, which was found to be inversely related to vegetable intake among women in Greece¹⁰⁰ and to total vitamin A intake, largely from vegetable sources, among women in New York State¹⁰¹. This possibility requires further examination.

Alcohol

Rapidly accumulating evidence, summarized elsewhere¹⁰², indicates that intake of alcoholic beverages is associated with a higher risk of breast cancer (Fig. 2). Positive associations with beer, wine and liquor in various populations indicates that the effect is due to alcohol *per se*¹⁰³. The possibility cannot be entirely excluded that some other characteristic of women who drink alcohol might explain this association, but such a characteristic would have to cause a larger attributable fraction of breast cancers than the known risk factors and also be strongly associated with alcohol use. When risk factors for breast cancer were allowed for, the association with alcohol was not materially altered in any of the published studies. If the association is truly one of cause and effect, possible interactions with other risk factors may be important. In one case-control study¹⁰⁴, only alcohol consumed before the age of 30 years contributed to increased risk. If confirmed in other studies, this would mean that if a middle-aged woman reduced her alcohol consumption,

it would not alter her risk of breast cancer.

In a recent cohort study there were higher rates of colon cancer among regular drinkers, especially women, than among non-drinkers¹⁰⁵. Further prospective data are needed that clearly distinguish colon from rectal cancers, for which the data implicating alcohol are somewhat stronger. A weak positive association, reviewed elsewhere¹⁰⁵, between use of alcoholic beverages and risk of colon cancer has been observed sporadically in case-control studies.

Other specific dietary factors

Many other dietary factors have been postulated as either protective or causative for breast and colon cancer. Putative protective factors include selenium¹⁰⁶; vitamins C¹⁰⁷ and E¹⁰⁸; preformed vitamin A and its precursor, β -carotene¹⁰⁹; vitamin D⁷⁸; naturally occurring inhibitors in plants such as the indoles⁹⁸; protease inhibitors⁹⁷ and fish oils¹¹⁰. Other causative factors suggested include specific fatty acids, such as the *trans*-isomers produced in processing vegetable oils¹¹¹, and mutagenic products formed in the cooking of food¹¹². Calcium may reduce risk of colon cancer¹¹³; reduced colonic epithelial-cell proliferation has been reported in subjects supplemented with calcium carbonate¹¹⁴ and there is some epidemiological evidence^{78,115}. Data are at present too limited either to refute or to confirm any of these hypotheses with confidence.

Energy balance

In animals, restricting energy intake reduces the occurrence of mammary and colon tumours and other neoplasms^{60-62,116,117}; accumulating evidence suggests that reduced energy intake, particularly early in life, may also reduce the incidence of breast cancer in humans. The influence of energy intake on cancer risk in humans is best studied using physical measurements that reflect energy balance at different periods in life. Substantial restriction of energy intake before the age of 20 will stunt height. Among adults, weight standardized for height and measures of change in weight are sensitive indicators of long-term energy balance.

Internationally, the high correlation (0.8) between the average height of adult women in a country and national rates of breast cancer¹¹⁸ provides some evidence for a protective effect of energy restriction before adulthood. Positive associations have been found between height and risk of breast cancer in several countries that experienced periods of food shortage, such as Greece¹¹⁹, the Netherlands¹²⁰ and Japan¹²¹. In the United

States¹²² and Scandinavia¹²³, height is generally unrelated to breast cancer, compatible with a sufficient supply of energy for most individuals to attain their genetically determined maximum height. In a recent sample of women from the United States, that was enriched in individuals at increased risk of malnutrition, a positive association between height and risk of breast cancer was found¹²⁴.

The relationship between obesity and risk of breast cancer is complex. In a large prospective study in the United States, mortality from breast cancer was about 50% higher among women who were more than 40% above average weight for their height than among women of average weight¹²⁵. This may partly be because tumours are diagnosed later among obese women¹²², rather than because of an aetiological role of obesity. In studies that considered the incidence of breast cancer rather than mortality, a distinct difference has been found between pre- and post-menopausal women. Before menopause, greater relative weight is associated with a reduced risk of breast cancer^{122,126}; rates among those in the heaviest 20% are about 40% lower than among the thinnest. Among post-menopausal women, positive associations with relative weight have been seen in several studies¹²⁷⁻¹³¹, but these have been generally weak, without a clear dose-response relationship, and in some studies no association was observed^{126,132}.

A weak positive association was seen between relative weight and mortality from colon cancer in men, but not women, in one large prospective study¹²⁵. However, in some case-control⁶⁸ and prospective cohort^{133,134} studies, relative weight had little association with incidence of colon cancer.

Contrasts between breast and colon cancers

Although the striking international correlations of both breast and colon cancer rates with indices of affluence suggest common aetiological factors, other evidence indicates important differences in aetiology or the period of life during which the factors act. For example in the United States, Japanese immigrants have rates of colon cancer similar to those of the caucasian population, whereas rates of breast cancer take several generations to approach those of the resident population^{135,136}. Rates of breast cancer among Seventh-Day Adventists are similar to those in the general population, but rates of colon cancer are about 40% lower⁵⁸.

The relationship between changes in diet and rates of mortality from breast and colon cancer are uniquely illustrated in Japan, where there has been a dramatic change in diet since World War II (Fig. 3). Between 1955 and 1975, fat intake increased from about 10% to about 25% of total energy intake. This has been accompanied by a striking rise in mortality from colon cancer but only a small change in breast cancer mortality. These data do not necessarily indicate that the fat composition of the diet caused the rapid rise in colon cancer, but they do suggest strongly that environmental changes will more rapidly influence rates of colon than breast cancer and that large changes in fat intake will be accompanied by changes in rates of breast cancer that are, at most, modest or substantially delayed.

It is useful to compare case-control and prospective cohort studies of colon and breast cancer conducted by the same investigators using similar methods, because the results are relatively unobscured by the differences typically found in study populations and research designs. Such pairs of studies have been conducted in upstate New York^{96,101}, Greece^{93,100}, Canada^{68,40}, Californian Seventh-Day Adventists⁴¹, Australia^{69,37} and nurses in the United States^{47,79}. Except in the upstate New York study, clear associations were seen between fat or meat intake and risk of colon cancer, but not of breast cancer, in each pair of analyses. These comparisons provide further evidence that the failure to find associations between diet and breast cancer in case-control and cohort studies is not simply due to lack of heterogeneity in diet or imprecise methods of assessing diet.

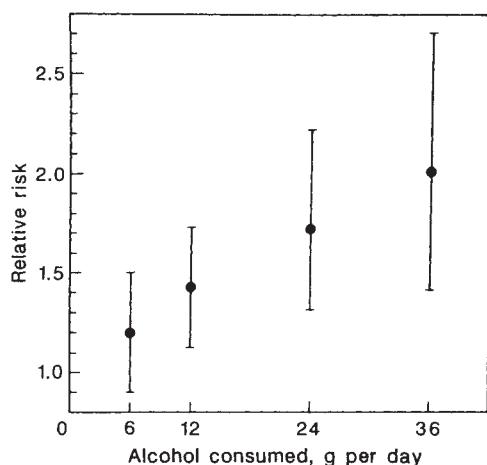


FIG. 2 Risk of breast cancer related to level of alcohol consumption, with 95% confidence intervals, based on pooled estimates from five prospective studies¹⁰². In women consuming 12 g of alcohol per day, about one standard drink of beer, wine or liquor, the risk of breast cancer was ~40% higher than in non-drinkers. The risk for 24 g per day rose to ~70% and for 36 g per day to ~100% compared with non-drinking women. (From ref. 111, with permission of the American Medical Association.)

Prospects

In spite of strong evidence that non-genetic factors account for the high rates of breast and colon cancer in industrialized countries, few specific aetiological factors are well established that offer the potential for prevention. Among the associations considered here, drinking alcohol and the risk of breast cancer is the relationship most strongly supported by epidemiological findings, but it is not certain whether this represents cause and effect. The benefits of reducing alcohol intake remain unclear because any influence on breast cancer risk may be limited to young adulthood. In adult women, a modest reduction in dietary fat and avoiding obesity seem likely to have little, if any, impact on reducing the risk of breast cancer. On the contrary, accumulating data indicate that being thin before menopause may actually increase the risk of breast cancer.

Several lines of evidence indicate that risk of colon cancer may be more responsive than breast cancer to changes in diet. Associations between dietary fat and colon cancer have been found in several studies; whether this represents a causal effect of the fat composition of the diet remains unproven and the type of fat rather than its total amount may be most important. Inverse associations have been found repeatedly between consumption of fruits and vegetables and the risk of colon cancer. The process of translating these epidemiological findings into public recommendations has taken an interesting turn: data on fruit and vegetable consumption have been used to compute fibre intake and, based on the resulting inverse associations with fibre intake, recommendations have been made to consume more whole-grain cereal, even though there is a consistent lack of evidence for a protective role of cereal fibre in epidemiological studies.

Except for limiting alcohol intake, which could at most only modestly decrease the incidence of breast cancer, prospects for prevention of breast cancer seem limited in the near future. The potential for preventing colon cancer by changes in diet seems greater, but exactly how needs to be more clearly defined. Increasing the intake of fruits and vegetables seems promising and reducing consumption of animal fats may prove to be effective. Large prospective studies of diet that are now under-

way or planned, together with randomized trials of dietary factors in relation to recurrence of colon polyps that are now in progress, will provide much additional information about the effects of specific foods and nutrients. □

Walter Willett is at the Harvard School of Public Health, Departments of Epidemiology and Nutrition, and the Channing Laboratory, Department of Medicine, Harvard Medical School and Brigham and Women's Hospital. Address correspondence to Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA

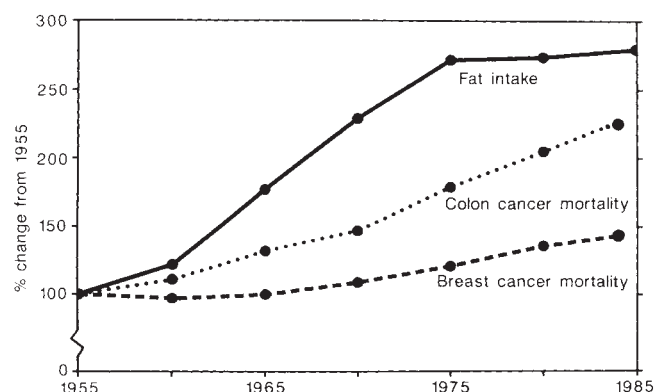


FIG. 3 Relative changes in fat intake and age-adjusted rates of mortality from breast and colon cancer in Japan between 1955 and 1985. During this period, fat intake increased by ~180% (from ~10–25% of energy intake) whereas total energy intake remained nearly constant. By 1984, age-adjusted mortality from breast cancer had increased ~45%, whereas mortality from colon cancer among women increased by 130%, an increase proportional to increased fat intake with a lag of ~15 years. Much of the modest increase in breast cancer mortality is probably a result of marked changes in reproductive patterns; for example, the percentage of women having their first child before the age of 19 decreased from 13 to 4% between 1950 and 1975¹⁰. Although data for incidence of breast cancer are available for a more limited period, they also indicate only a slow increase over time¹³⁵. Absolute levels in 1955 were 20.3 g per day of fat per capita, 1.9 colon cancer deaths per 10⁵ person-years, and 3.3 breast cancer deaths per 10⁵ person-years. (Based on data from Health and Welfare Statistics Association, Japan¹¹.)

- Horm, J. W., Asire, A. J., Young, J. L. & Pollack, E. S. *NIH Pub. No. 85-1837* (1984).
- Anderson, D. E. *Cancer* **34**, 1090–1097 (1974).
- Bain, C., Speizer, F. E., Rosner, B., Belanger, C. & Hennekens, C. H. *Am. J. Epidemiol.* **111**, 301–308 (1980).
- Macklin, M. T. *J. natn. Cancer Inst.* **24**, 551–571 (1960).
- Cannon-Albright, L. A., Skolnick, M. H., Bishop, D. T., Lee, R. G. & Burt, R. W. *New Engl. J. Med.* **319**, 533–537 (1988).
- Armstrong, B. & Doll, R. *Int. J. Cancer* **15**, 617–631 (1975).
- McMichael, A. J. & Giles, G. G. *Cancer Res.* **48**, 751–756 (1988).
- Haenszel, W. & Kurihara, M. *J. natn. Cancer Inst.* **40**, 43–68 (1968).
- Bjarnason, O., Day, N., Snaedal, G. & Tuulius, H. *Int. J. Cancer* **13**, 689–696 (1974).
- Health and Welfare Statistics Ass. (Japan) *Indices of Health and Welfare: Trends of National Health* **33**, 9:54 (1986).
- Doll, R. & Peto, R. *J. natn. Cancer Inst.* **66**, 1191–1308 (1981).
- Mustacchi, P. *Arch. int. Med.* **108**, 195–198 (1961).
- MacMahon, B. et al. *Bull. Wild Hlth Org.* **42**, 185–194 (1970).
- Kampert, J. B., Whittemore, A. S. & Paffenbarger, R. S. *Am. J. Epidemiol.* **128**, 962–979 (1988).
- Choi, N. W. et al. *Am. J. Epidemiol.* **107**, 510–521 (1978).
- Cairns, J. *Nature* **255**, 197–200 (1975).
- Russo, J., Tay, L. K. & Russo, I. H. *Breast Cancer Res. Treat.* **2**, 5–73 (1982).
- Kelsey, J. L. *Epidemiol. Rev.* **1**, 74–109 (1979).
- MacMahon, B. et al. *Int. J. Cancer* **29**, 13–16 (1986).
- Trichopoulos, D., MacMahon, B. & Cole, P. *J. natn. Cancer Inst.* **48**, 605–613 (1972).
- MacMahon, B., Cole, P. & Brown, J. *J. natn. Cancer Inst.* **50**, 21–42 (1973).
- Pike, M. C. et al. *Nature* **303**, 767–770 (1983).
- Key, T. J. & Pike, M. C. *Eur. J. Cancer Clin. Oncol.* **24**, 29–43 (1988).
- Prentice, R. L. & Thomas, D. B. *Adv. Cancer Res.* **49**, 285–401 (1987).
- Miller, D. R. *Am. J. Epidemiol.* **129**, 269–280 (1989).
- Armstrong, B. K. *Med. J. Austral.* **148**, 213–214 (1988).
- Brinton, L. A., Hoover, R. & Fraumeni, J. F. Jr. *Br. J. Cancer* **54**, 825–832 (1986).
- McGregor, D. H. *J. natn. Cancer Inst.* **59**, 799–811 (1977).
- Boice, J. D. & Monson, R. R. *J. natn. Cancer Inst.* **59**, 823–832 (1977).
- Dupont, W. D. & Page, D. L. *New Engl. J. Med.* **312**, 146–151 (1985).
- Schottenfeld, D. & Winer, S. J. in *Cancer Epidemiology and Prevention* (eds Schottenfeld, D. & Fraumeni, J. F.) 703–727 (Saunders, Philadelphia, 1982).
- Muto, T., Bussey, H. J. & Morson, B. C. *Cancer* **36**, 2251–2270 (1975).
- Diet, Nutrition, and Cancer (Committee on Diet, Nutrition and Cancer; National Research Council) (National Academy Press, Washington DC, 1982).
- Chen, J., Campbell, T. C., Junyao, L. & Peto, R. *The Dietary, Lifestyles, and Mortality Characteristics of 65 Rural Populations in the Peoples Republic of China* (Division of Nutritional Sciences, Cornell Univ., Cornell, 1987).
- Graham, S. et al. *Am. J. Epidemiol.* **116**, 68–75 (1982).
- Hirohata, T., Nomura, A. M. Y., Hankin, J. H., Kolonel, L. N. & Lee, J. *J. natn. Cancer Inst.* **78**, 595–600 (1987).
- Rohan, T. E., McMichael, A. J. & Baghurst, P. A. *Am. J. Epidemiol.* **128**, 478–489 (1988).
- Katsouyanni, K. et al. *Cancer* **61**, 181–185 (1988).
- Hirohata, T. et al. *Natn. Cancer Inst. Monogr.* **69**, 187–190 (1985).
- Miller, A. B. et al. *Am. J. Epidemiol.* **107**, 499–509 (1978).
- Phillips, R. L. *Cancer Res.* **35**, 3513–3522 (1975).
- Talamini, R. et al. *Br. J. Cancer* **49**, 723–739 (1984).
- Le, M. G., Moulton, L. H., Hill, C. & Kramar, A. J. *J. natn. Cancer Inst.* **77**, 633–636 (1986).
- Hislop, T. G., Coldman, A. J., Elwood, J. M., Brauer, G. & Kan, L. *Cancer Det. Prev.* **9**, 47–58 (1986).
- Lubin, J. H. et al. *Int. J. Cancer* **28**, 685–689 (1982).
- Lubin, J. H., Wax, Y. & Modan, B. *J. natn. Cancer Inst.* **77**, 605–612 (1986).
- Willett, W. C. et al. *New Engl. J. Med.* **316**, 22–28 (1987).
- Jones, D. Y. et al. *J. natn. Cancer Inst.* **79**, 465–471 (1987).
- Hirayama, T. *Prev. Med.* **7**, 173–195 (1978).
- Phillips, R. L. & Snowdon, D. A. *Cancer Res.* **43**, 2403–2408 (1983).
- Mills, P. K., Annegers, J. F. & Phillips, R. L. *Am. J. Epidemiol.* **127**, 440–453 (1988).
- Prentice, R. L. et al. *J. natn. Cancer Inst.* **80**, 802–814 (1988).
- Goodwin, P. J. & Boyd, N. F. *J. natn. Cancer Inst.* **79**, 473–485 (1987).
- Rosner, B. A., Willett, W. C. & Spiegelman, D. *Statistics in Medicine*, in the press.
- Rosner, B. A., Willett, W. C. & Spiegelman, D. *Am. J. Epidemiol.* **128**, 906 (1988).
- National Cancer Institute, Washington DC, D.H.H.S. Publ. no. (NIH) 84-2671 (1984).
- Kinlen, L. J. *Lancet* **1**, 946–949 (1982).
- Phillips, R. L. et al. *J. natn. Cancer Inst.* **65**, 1097–1107 (1980).
- Birt, D. F. *Adv. exp. Med. Biol.* **206**, 69–84 (Plenum, New York, 1986).
- Pariza, M. W. & Simopoulos, A. P. *Am. J. Clin. Nutr.* **45**, 149–272 (1987).
- Boissonneault, G. A., Elson, C. E. & Pariza, M. W. *J. natn. Cancer Inst.* **76**, 335–338 (1986).
- Albanes, D. *Cancer Res.* **47**, 1887–1892 (1987).
- Appleton, B. S. & Landers, R. E. in *Essential Nutrients in Carcinogenesis* (eds Poirer, L. A., Newburne, P. M. & Pariza, M. W.) *Adv. exp. Med. Biol.* **206**, 99–104 (1986).
- Sonnenschein, E., Glickman, L., McKee, L. & Goldschmidt, M. *Am. J. Epidemiol.* **126**, 736 (1987).
- Willett, W. C. & MacMahon, B. *New Engl. J. Med.* **310**, 697–703 (1984).
- Hagerly, M. A., Howie, B., Tan, S. & Shultz, T. D. *Fedn Proc.* **46**, 587 (1987).
- Rose, D. P., Boyar, A. P., Cohen, C. & Strong, L. E. *J. natn. Cancer Inst.* **78**, 623–626 (1987).
- Jain, M. et al. *Int. J. Cancer* **26**, 757–768 (1980).
- Potter, J. D. & McMichael, A. J. *J. natn. Cancer Inst.* **76**, 557–569 (1986).
- Lyons, J. L. et al. *J. natn. Cancer Inst.* **78**, 853–861 (1987).
- Bristol, J. B., Emmett, P. M., Heaton, K. W. & Williamson, R. C. *Br. Med. J.* **291**, 1467–1470 (1985).
- Graham, S. et al. *Am. J. Epidemiol.* **128**, 490–503 (1988).
- Kune, G. A., Kune, S. & Watson, L. F. *Nutr. Cancer* **9**, 1–56 (1987).
- Macquart-Moulin, G. et al. *Int. J. Cancer* **38**, 183–191 (1986).

75. Berta, J. L., Coste, T., Rautouseau, J., Guilloud-Bataille, M. & Pequignot, G. *Gastroenterol. Clin. Biol.* **9**, 348–353 (1985).
76. Tuyns, A. J., Haeletermans, M. & Kaaks, R. *Nutr. Cancer* **10**, 181–196 (1987).
77. Stemmermann, G. N., Nomura, A. M. Y. & Heilbrun, L. K. *Cancer Res.* **44**, 4633–4637 (1984).
78. Gariand, C. *et al. Lancet* **i**, 307–309 (1985).
79. Stampfer, M. J. *et al. Fedn Proc.* **46**, 3303 (1987).
80. Morgan, J. W., Fraser, G. E., Phillips, R. L. & Andress, M. H. *Am. J. Epidemiol.* **128**, 918 (1988).
81. Garabrant, D. H. *et al. Am. J. Epidemiol.* **119**, 1,005–1,014 (1984).
82. Slattery, M. L., Schumacher, M. C., Smith, K. R., West, D. W. & Abd-Elghany, N. *Am. J. Epidemiol.* **128**, 989–999 (1988).
83. Willett, W. C. & Stampfer, M. J. *Am. J. Epidemiol.* **124**, 17–27 (1986).
84. Howe, G. R., Miller, A. B. & Jain, M. *Am. J. Epidemiol.* **124**, 157–159 (1986).
85. Reddy, B. S. *Cancer Res.* **41**, 3766–3768 (1981).
86. Goldin, B. R., Swenson, L., Dwyer, J., Sexton, M. & Gorbach, S. L. *J. natn. Cancer Res.* **64**, 255–261 (1980).
87. Burkitt, D. P. *Cancer* **28**, 3–13 (1971).
88. Weisburger, J. H. & Wynder, E. L. *Important Advances in Oncology* (eds DeVita, V. T., Hellman, S. & Rosenberg, S. A.) 197–270 (J. B. Lippincott, Philadelphia, 1987).
89. McKeown-Eyssen, G. E. & Bright-See, E. *Nutr. Cancer* **6**, 160–170 (1984).
90. Modan, G. *et al. J. natn. Cancer Inst.* **61**, 709–714 (1978).
91. Miller, A. B., Howe, G. R., Jain, M., Craib, K. J. & Harrison, L. *Int. J. Cancer* **32**, 155–161 (1983).
92. LaVecchia, C. *et al. Int. J. Cancer* **41**, 492–498 (1988).
93. Manousos, O. *et al. Int. J. Cancer* **32**, 1–5 (1983).
94. Martinez, I., Torres, R., Frias, Z., Colon, J. R. & Fernandez, N. *Adv. med. Oncol. Res. Education* **3** (ed. Birch, J. M.) 45–52 (Pergamon, New York, 1979).
95. Haenszel, W., Locke, F. B., & Segi, M. *J. natn. Cancer Inst.* **64**, 17–22 (1980).
96. Graham, S., Dayal, H., Swanson, M., Mittelman, A. & Wilkinson, G. *J. natn. Cancer Inst.* **61**, 709–714 (1975).
97. Troll, W., Wiesner, R. & Frenkel, K. *Adv. Cancer Res.* **49**, 265–283 (1987).
98. Wattenberg, L. W. & Loub, W. D. *Cancer Res.* **38**, 1,410–1,413 (1978).
99. Modan, B., Cuckle, H. & Lubin, F. *Int. J. Cancer* **28**, 421–424 (1981).
100. Katsouyanni, K. *et al. Int. J. Cancer* **38**, 185–200 (1986).
101. Graham, S. *et al. Am. J. Epidemiol.* **116**, 68–75 (1982).
102. Longnecker, M. P., Berlin, J. A., Orza, M. J. & Chalmers, T. C. *J. Am. med. Ass.* **260**, 652–656 (1988).
103. Willett, W. C. *et al. New Engl. J. Med.* **314**, 1174–1180 (1987).
104. Harvey, E. B., Schairer, C., Brinton, L. A., Hoover, R. N. & Fraumeni, J. F., Jr. *J. natn. Cancer Inst.* **78**, 657–661 (1987).
105. Klatzky, A. L., Armstrong, M. A., Friedman, G. D. & Hiatt, R. A. *Am. J. Epidemiol.* **128**, 1007–1015 (1988).
106. Stampfer, M. J., Colditz, G. A. & Willett, W. C. *Cancer Surveys* **6**, 623–633 (1987).
107. Block, G. & Merkes, M. *Nutrition and Cancer Prevention* (eds Moon, T. E. & Micozzi, M. S.) 341–388 (Dekker, New York, 1988).
108. Mergens, W. J. & Bhagavan, H. N. *Nutrition and Cancer Prevention* (eds Moon, T. E. & Micozzi, M. S.) 305–340 (Dekker, New York, 1988).
109. Peto, R., Doll, R., Buckley, J. D. & Sporn, M. D. *Nature* **290**, 201–208 (1981).
110. Nelson, R. L., Tanure, J. C., Andrianopoulos, G., Souza, G. & Lands, W. E. M. *Nutr. Cancer* **11**, 215–220 (1988).
111. Enig, M. G., Munn, R. J. & Keeney, M. *Fedn Proc.* **37**, 2215–2220 (1978).
112. Lyon, J. L. & Mahoney, A. W. *Am. J. Epidemiol.* **128**, 1000–1006 (1988).
113. Newmark, H. L., Wargovich, M. J. & Bruce, W. R. *J. natn. Cancer Inst.* **72**, 1323–1325 (1984).
114. Lipkin, M. & Newmark, H. *New Engl. J. Med.* **313**, 1381–1384 (1985).
115. Sorenson, A. W., Slattery, M. L. & Ford, M. H. *Nutr. Cancer* **11**, 135–145 (1988).
116. Kritchevsky, D. & Klurfeld, D. M. *Adv. exp. Med. and Biol.* **206**, 55–68 (1986).
117. Weindruch, J. R. & Walford, R. L. *Science* **215**, 1415–1418 (1982).
118. Micozzi, M. S. *Yearbook Phys. Anthropol.* **28**, 175–206 (1985).
119. Valaoras, V. G. *et al. Int. J. Cancer* **4**, 350–363 (1969).
120. de Waard, F. *Cancer Res.* **35**, 3351–3356 (1975).
121. Hirayama, T. *Prev. Med.* **7**, 173–195 (1978).
122. Willett, W. C. *et al. Am. J. Epidemiol.* **122**, 731–740 (1985).
123. Adami, H. O. *et al. Br. J. Cancer* **36**, 787–792 (1977).
124. Swanson, C. A., Jones, Y. D., Schatzkin, A., Brinton, L. A. & Ziegler, R. G. *Cancer Res.* **48**, 5363–5367 (1988).
125. Lew, E. A. & Garfinkel, L. *J. chronic Dis.* **32**, 563–576 (1979).
126. Le Marchand, L., Kolonel, L. N., Earle, M. E. & Mi, M. R. *Am. J. Epidemiol.* **128**, 137–152 (1988).
127. Heimich, S. P. *et al. Am. J. Epidemiol.* **117**, 35–45 (1983).
128. Paffenbarger, R. S., Kampert, J. B. & Chang, H.-G. *Am. J. Epidemiol.* **112**, 258–268 (1980).
129. Choi, N. W. *et al. Am. J. Epidemiol.* **107**, 510–521 (1978).
130. Lubin, F., Ruder, A. M., Wax, Y. & Modan, B. *Am. J. Epidemiol.* **122**, 579–588 (1986).
131. de Waard, F., Cornelius, J. P., Aoki, K. & Yoshida, M. *Cancer* **40**, 1269–1275 (1977).
132. London, S. *J. Am. J. Epidemiol.* **128**, 914 (abstract) (1988).
133. Sidney, S., Friedman, G. D. & Hiatt, R. A. *Am. J. Epidemiol.* **124**, 33–38 (1986).
134. Nomura, A., Heilbrun, L. K. & Stemmermann, G. N. *J. natn. Cancer Inst.* **74**, 319–323 (1985).
135. Hana, A. & Fujimoto, I. *Natn. Cancer Inst. Monogr.* **62**, 3–7 (1982).
136. Haenszel, W. *Cancer Epidemiology and Prevention* (eds Schottenfeld, D. & Fraumeni, T. F.) 194–207 (W. B. Saunders, Philadelphia, 1982).

ACKNOWLEDGEMENTS. I thank Drs Meir Stampfer, John Cairns, and Brian MacMahon for helpful comments and Doreen Hurd for assistance in the preparation of the manuscript.

ARTICLES

The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis

Daniel Finley*, Bonnie Bartel & Alexander Varshavsky

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Three of the four yeast ubiquitin genes encode hybrid proteins which are cleaved to yield ubiquitin and previously unidentified ribosomal proteins. The transient association between ubiquitin and these proteins promotes their incorporation into nascent ribosomes and is required for efficient ribosome biogenesis. These results suggest a novel 'chaperone' function for ubiquitin, in which its covalent association with other proteins promotes the formation of specific cellular structures.

THE covalent ligation of ubiquitin to various acceptor proteins in eukaryotic cells participates in or regulates a number of cellular processes, such as selective protein degradation^{1–3}, DNA repair⁴, progression through the cell cycle^{2,5,6} and a variety of stress responses^{7–11}. A major function of ubiquitin is to mark

proteins destined for selective elimination^{12–14}. An alternative role for ubiquitin has also been suggested¹⁵, in which it reversibly joins to an acceptor protein to modulate protein function without metabolically destabilizing the acceptor protein. *In vivo* acceptors of ubiquitin include histones H2A and H2B¹⁶, actin¹⁷, the lymphocyte homing receptor¹⁸, the platelet-derived growth factor receptor¹⁹, the growth hormone receptor²⁰ and the intracellular neurofibrillary tangles in neurodegenerative diseases such as Alzheimer's^{21–22}.

The conjugation of ubiquitin to other proteins involves the formation of an isopeptide bond between the C-terminal glycine residue of ubiquitin and the ε-amino group of a lysine residue in the acceptor protein^{1,12}. In the yeast *Saccharomyces cerevisiae*, the six or more enzymes that catalyse such reactions¹⁵ include the products of the genes *RAD6*, which is required for DNA repair⁴ and *CDC34*, which is required for the transition from the G1 to the S phase of the cell cycle⁵. Ubiquitinated proteins often contain single ubiquitin moieties linked to one or more of the acceptor protein's lysine residues^{12,16,17,23}. Alternatively, several ubiquitin moieties can be attached sequentially to an acceptor protein to form a chain of branched ubiquitin-ubiquitin conjugates in which the C-terminal Gly 76 of one ubiquitin is

* Present address: Department of Cellular and Molecular Physiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

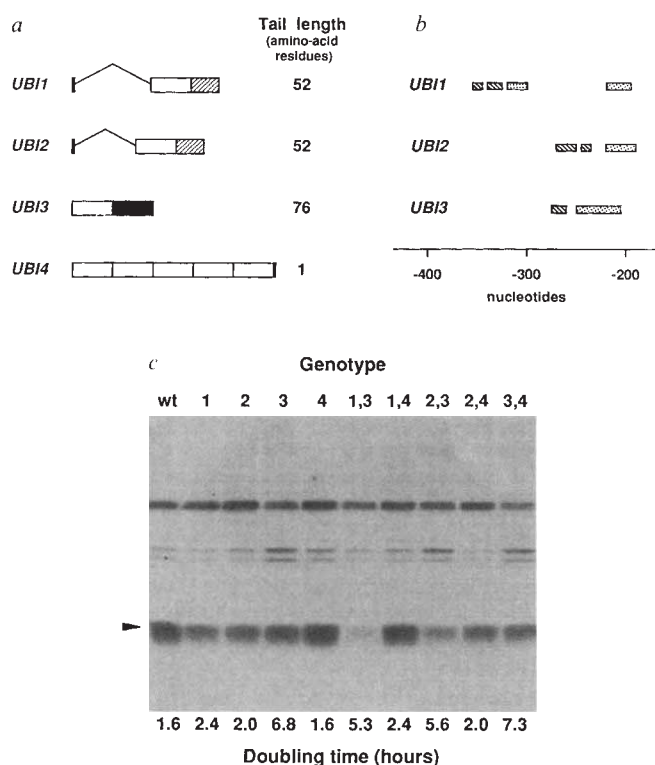


FIG. 1. *S. cerevisiae* ubiquitin genes and ubiquitin mutants. *a*, Organization of yeast ubiquitin genes. Open blocks represent the 228-base pair (bp) ubiquitin-coding repeats. Striped and shaded blocks represent tail-coding elements. *UBI1* and *UBI2* contain identically positioned, mostly nonhomologous introns within their ubiquitin-coding elements²⁴. *b*, Putative upstream activation (UAS) sites of *UBI1-UBI3* genes. Positions are relative to the ATG start codon in each gene. Striped blocks represent near-matches (11–14 out of 15) to the consensus sequence of UAS sites within genes for ribosomal proteins³⁰. Stippled blocks represent T-rich stretches often found downstream of such UAS sites³⁰. *c*, Immunoblot analysis of intracellular ubiquitin levels and doubling times of yeast strains bearing deletions of various ubiquitin genes. Wild-type lane is on the left; other lanes are numbered at the top according to the ubiquitin gene or genes deleted in the corresponding strain (for example, lane 1, *ubi1* strain). The ubiquitin band is indicated by an arrow. The doubling time of each strain at 30 °C in YPD, a rich medium⁴⁶, is indicated underneath; the same growth conditions were used in the immunoblotting experiment. The bands of high molecular weight presumably correspond to ubiquitin-protein conjugates.

METHODS. Exponentially growing cells were centrifuged at 5,000 *g* for 5 min at 4 °C, resuspended in 30 mM Tris-HCl (pH 6.8), 1 mM MgCl₂, centrifuged again, and resuspended in a buffer containing 30 mM Tris-HCl (pH 6.8), 1 mM MgCl₂, 2 mM *N*-ethylmaleimide, 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ leupeptin, 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ chymostatin and 10 µg ml⁻¹ pepstatin (Sigma). Cells were lysed by vortexing in the presence of glass beads, and the extracts were centrifuged at 12,000 *g* for 10 min. Protein (150 µg) was loaded onto each lane of an 18% polyacrylamide-SDS gel. Fractionated proteins were then electroblotted onto PVDF membranes (Millipore) for 20 min at 13 V cm⁻¹ (under these conditions, the transfer of high molecular weight proteins is not complete). Filters were probed with an antibody to ubiquitin²¹, followed by ¹²⁵I-labelled Protein A (ICN Radiochemicals) using standard techniques⁸. The filter-bound ¹²⁵I-Protein A was detected by autoradiography. Doubling times were determined by monitoring the absorbance at 600 nm of exponentially growing liquid cultures. All strains analysed were *MATα*. All strains used in this work are congenic, being derived from DF5 (*MATα/MATα*, *trp1-1 (am)/trp1-1 (am)*, *ura3-52/ura3-52*, *his3-Δ200/his3-Δ200*, *leu2-3, 2-112/leu2-3, 2-112*, *lys2-801/lys2-801*), a previously described homozygous diploid⁸. Precise chromosomal deletions were constructed essentially as described⁸, except that in the case of the *UBI2* *in vitro* deletion, we used restriction endonucleases rather than an oligonucleotide. A 1.6 kilobase (kb) genomic *HincII* fragment containing the *UBI2* gene²⁴ was subcloned into pUC13 (ref. 48), and the resulting plasmid was digested with *HindIII* and *Clal*, deleting all but 8 bp of the *UBI2* coding sequence and no noncoding sequence. Subsequent steps were as described⁸. The deletion-disruption mutation *ubi1* was marked with the *TRP1* gene, whereas *ubi2* was marked with *URA3* and *ubi3* with *HIS3*.

joined to the internal Lys 48 of an adjacent ubiquitin²³. The attachment of such a multi-ubiquitin chain to the proteolytic substrate is apparently essential for the degradation of a variety of proteins²³.

Unlike the branched ubiquitin-protein conjugates, which are formed post-translationally, linear ubiquitin adducts are formed as the translational products of natural gene fusions^{7,24,25}. Thus, in *S. cerevisiae* for example, ubiquitin is generated exclusively by proteolytic processing of precursors in which ubiquitin is joined either to itself, as in the linear polyubiquitin protein *UBI4*, or to unrelated 'tail' amino-acid sequences, as in the hybrid proteins *UBI1*, *UBI2* and *UBI3* (Fig. 1*a*). In growing yeast cells, most ubiquitin is generated from the *UBI1-UBI3* hybrid proteins⁸. The polyubiquitin (*UBI4*) gene is dispensable in growing cells but becomes essential (as the main supplier of ubiquitin) during stress⁸.

The tail amino-acid sequences of the ubiquitin precursors are highly conserved between yeast and mammals^{7,24,25}, suggesting that they function similarly in all eukaryotes. The yeast *UBI1* and *UBI2* genes encode identical 52-residue tails, whereas *UBI3* encodes a 76-residue tail with little amino-acid sequence similarity to the tails of the *UBI1* and *UBI2* proteins²⁴ (Fig. 1*a*). Both the *UBI1/UBI2* and *UBI3* tails are highly basic however, and contain cysteine-rich, putative zinc-binding domains, suggesting that these proteins may bind nucleic acids²⁴.

We report here that these tails are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. These and related results suggest a novel, 'chaperone' function for ubiquitin, in which its covalent association with other proteins can promote the formation of specific cellular structures.

Deletion analysis of the *UBI1-UBI3* genes

To investigate the functions of the ubiquitin-hybrid genes, we precisely deleted the coding sequences of these genes from the *S. cerevisiae* genome (see Fig. 1 legend). Unlike the deletion of *UBI4*⁸, deletions of the ubiquitin-hybrid genes resulted in slow growth phenotypes, with the *ubi3* deletion having a much more severe effect than the *ubi1* or *ubi2* deletions (Fig. 1*c*). The phenotypes of the *ubi1* and *ubi2* single mutants can be viewed essentially as gene dosage effects, because *UBI1* and *UBI2* encode identical proteins²⁴ and because the *UBI2* gene, carried on a centromere-containing (*CEN*) plasmid, fully complemented the growth defect of the *ubi1* mutant (data not shown). The *ubi1 ubi2* double mutant is not viable, as indicated by our inability to recover *ubi1 ubi2* meiotic segregants upon sporulation of a *ubi1/+*, *ubi2/+* diploid, and by the fact that a yeast plasmid carrying the *UBI2* gene and an origin of replication, but lacking a *CEN* element was fully stabilized against mitotic loss in a *ubi1 ubi2* genetic background (data not shown). Thus, *UBI1* and *UBI2* encode an essential protein and seem to be functionally interchangeable genes.

The relative contributions of *UBI1-UBI4* to steady-state levels of free ubiquitin were assessed by immunoblotting. Deletion of any one of the four yeast ubiquitin genes resulted in, at most, a slight decrease in the level of free ubiquitin (Fig. 1*c*). Strikingly, there was no correlation between growth rates of the various deletion mutants and their levels of free ubiquitin (Fig. 1*c*). The growth defects of the *ubi1*, *ubi2* and *ubi3* deletion mutants must therefore be due, at least in part, to either the absence or decrease in concentration of specific tail components of the *UBI1-UBI3* proteins.

To further dissect the contributions of ubiquitin and the tails to the functions of the ubiquitin-hybrid genes, genetic elements encoding the tail and ubiquitin moieties were separated from each other in plasmid constructs, introduced into the mutant strains described above, and tested for their ability to complement the cognate deletion mutations (Fig. 2). Oligonucleotide-directed mutagenesis was used to construct a plasmid in which the entire ubiquitin-coding component of *UBI3*, except for its initiator ATG codon, had been deleted. This plasmid, which

retained the flanking regions of the *UBI3* gene, but expressed the *UBI3* tail in a ubiquitin-free form, fully complemented the growth deficiency of the *ubi3* deletion mutant. Other *ubi3* phenotypes, such as hypersensitivity to nitrogen starvation, defective processing of RNA precursors (see below) and the abnormally large size of *ubi3* cells, were also complemented by the *UBI3* tail-expressing plasmid. Although in these experiments the *UBI3* tail gene was carried on a low copy number *CEN* plasmid, the observed full complementation of the *ubi3* phenotype depended on a moderate amplification of the plasmid copy number (see below). In contrast to the *UBI3* tail, free ubiquitin, when expressed from the *UBI3* promoter, did not complement any of the *ubi3* phenotypes significantly. A plasmid, pUb39, which expressed free ubiquitin from the *CUP1* promoter²⁶ at a rate comparable to that of *UBI1-UBI4* combined, also failed to complement *ubi3* (Fig. 2).

In analogous experiments with *UBI1*, we expressed ubiquitin and the *UBI1* tail separately using heterologous promoters, as plasmids carrying the *UBI1* promoter sequence were unstable in yeast. Expression of the free *UBI1* tail from the strong *GPD* promoter²⁷ complemented the lethal phenotype of the *ubi1 ubi2* double deletion and supported wild-type growth rates in this genetic background (Fig. 2). These results indicated that the *UBI1* and *UBI2* tails are essential proteins and can function without an N-terminal ubiquitin moiety. In a complementary experiment, the free ubiquitin-expressing plasmid pUb39 failed to rescue the viability (ability to germinate) of *ubi1 ubi2* spores (Fig. 2).

Vector	Promoter	Product	Complementation					
			<i>ubi1</i>	<i>ubi2</i>	<i>ubi3</i>	<i>ubi1 ubi2</i>	<i>ubi3</i>	
pUb94		UBI1 tail	+		ND		ND	
pUb36		UBI3 tail		ND	+		ND	
pUb95		UBI1 tail UBI3 tail	+		+	+	+	
pUb39		Ubiquitin	—		—		—	
pUb35		Ubiquitin	ND		—		ND	

FIG. 2 Summary of plasmid-based genetic complementation tests. Plasmids were transformed as described⁴⁶ into singly or multiply heterozygous diploids of the appropriate genotype. Transformants were sporulated and appropriate meiotic products of both mating types were tested for viability and growth rates. ND, not determined.

METHODS. Growth was assayed in media selective for the plasmid-borne selectable marker, wild-type growth being taken as that of a chromosomally wild-type strain transformed with the parental plasmid lacking ubiquitin genes and their derivatives. To construct pUb94, the ubiquitin-coding element of *UBI1* was deleted (except for its start codon) from a *UBI1*-containing DNA fragment, as described⁴⁷ (see also Fig. 1 legend). A 340-bp *HinfI* fragment containing the tail-coding element of *UBI1* was ligated to *Bam*HI linkers and subcloned into *Bam*HI-digested pGPD(G)-2 (ref. 27; pGPD(G)-2 has a UAS^{gal}-containing DNA fragment inserted into the *SalI* site of pGPD(s)-2; G. Bitter, personal communication) to yield pUb92. In this way the *GPD* promoter was positioned in place of the *UBI1* promoter. The expression cassette thus constructed was subcloned into the *SmaI* site of YEp351 (ref. 49), yielding pUb94. pUb36 is a YCp50-based⁵⁰ vector containing a *UBI3* tail-coding element (the ubiquitin-coding element is deleted), about 700 bp of 5' noncoding sequence from the *UBI3* gene, and about 400 bp of 3' noncoding sequence. pUb95 was constructed by subcloning the pUb36 insert into the unique *XbaI* site of pUb94. pUb39 is identical to YEp46 (ref. 26) except that it carries the *LYS2* rather than *TRP1* selectable marker. pUb35 is identical to pUb36 except that its insert contains the *UBI3* ubiquitin-coding element (the tail-coding element is deleted). Because the expression of the *UBI1* tail from the plasmids pUb94 and pUb95 is galactose-dependent, glucose was replaced in the medium by 2% galactose, 2% glycerol, 2% ethanol and 40 μ g ml⁻¹ aspartic acid⁵¹ when these plasmids were assayed for complementation.

These results indicate that the tails of ubiquitin-hybrid proteins have growth-related functions and that location of the tails within ubiquitin-hybrid proteins is not strictly essential for tail function. It has been proposed that the ubiquitin-tail fusion arrangement may instead be required for ubiquitin function²⁸; the tails, by virtue of their apparent nuclear targeting signals^{24,28}, might mediate transport of ubiquitin to the nucleus. To test this possibility, a strain entirely lacking ubiquitin-hybrid proteins was generated. We constructed a plasmid, pUb95, that expressed both *UBI1* and *UBI3* tails in a ubiquitin-free form and transformed it into a *ubi1/+*, *ubi2/+*, *ubi3/+* diploid, which was sporulated and subjected to tetrad analysis (Fig. 2). In the presence of the free tail-expressing plasmid, *ubi1 ubi2 ubi3* meiotic products could be recovered as viable cells, indicating that linkage of the tails to ubiquitin within the ubiquitin-hybrid proteins does not support an essential function of ubiquitin. The polyubiquitin gene, *UBI4*, the only ubiquitin gene that remained in the *ubi1 ubi2 ubi3* mutants, was strongly induced in this genetic background (unpublished data).

Deficiency of ribosomal subunits

In the course of a northern hybridization experiment, RNA preparations from the various *ubi1-ubi3* mutants were analysed electrophoretically. This revealed a strikingly substoichiometric ratio of 18S to 25S rRNA in the *ubi3* mutants (Fig. 3a). The 18S and 25S rRNAs reside within the small (40S) and large (60S) ribosomal subunits, respectively, and their normal 1:1 stoichiometry is maintained partly because they are derived from a single precursor transcript^{29,30}.

To test whether the deficiency of 18S rRNA in the *ubi3* mutants was accompanied by a deficiency of small ribosomal subunits, we prepared extracts from *ubi3* and wild-type cells under conditions which dissociate translating polyribosomes³¹. The extracts were sedimented through 15–30% sucrose gradients at low ionic strength, which promotes the association of large and small ribosomal subunits into 80S ribosomes (monosomes). The main sedimentation peak in the wild-type extract was monosomal (Fig. 4a). In contrast, the main peak in the *ubi3* extract was composed of free 60S subunits, and the monosome peak was greatly reduced. These results indicate that the *ubi3* mutants contain substoichiometric quantities of small ribosomal subunits.

To determine whether the *ubi3* mutant was defective in the formation (rather than metabolic stability) of 40S ribosomal subunits, we examined the processing of pre-rRNA in mutant and wild-type cells. Pre-rRNA processing was monitored using a pulse-chase procedure, followed by agarose gel electrophoresis and fluorography. The results are shown in Fig. 3c and d, together with a diagram of the pre-rRNA processing pathway (Fig. 3b). After a 6-min pulse, the predominantly labelled species are the 27S and 20S RNAs, which are the immediate precursors of the mature 25S and 18S rRNAs, respectively. In wild-type cells at 30 °C, these intermediate species are processed to completion within 30 min, whereas in the *ubi3* cells the rate at which the 18S rRNA is generated from the 20S pre-rRNA is dramatically reduced, and the bulk of the 20S pre-rRNA is ultimately degraded rather than processed. In contrast, the conversion of the 27S precursor to the 25S rRNA (residing in the large ribosomal subunit) is normal in the *ubi3* mutant (Fig. 3d). The 20S pre-rRNA processing defect cannot be accounted for by a defect in its transport to the cytoplasm (data not shown), where the 20S pre-rRNA is normally converted to 18S rRNA²⁹. Thus, a specific pre-rRNA processing defect explains the aberrant stoichiometry of rRNA in the *ubi3* mutant and accounts, at least in part, for its growth defect.

Remarkably, both the *ubi1 ubi3* and *ubi2 ubi3* double mutants grow faster than the *ubi3* single mutant, despite their relatively low free ubiquitin levels (Fig. 1c). This is particularly striking as the *ubi1* and *ubi2* single mutants grow more slowly than wild-type cells (Fig. 1c). Thus, *ubi1* and *ubi2* are both weak

suppressors of *ubi3*. Figure 3a shows that *ubi1* and *ubi2* also partially restore the stoichiometry of 18S rRNA to 25S rRNA in the *ubi3* genetic background. This result indicates that *UBI1* and *UBI2* are, like *UBI3*, involved in some aspect of ribosome assembly or turnover. Although the ratio of 18S rRNA to 25S rRNA, as estimated from gel electrophoretic patterns (Fig. 3a), is not noticeably abnormal in either *ubi1* and *ubi2* mutants, a more sensitive assay, namely sucrose gradient sedimentation, revealed a deficiency of 60S subunits in the *ubi1* mutant, in contrast to the *ubi3* mutant which was deficient in 40S subunits (Fig. 4b). A slight reduction in the monosome level (relative to total RNA in the extract) was also apparent in the *ubi1* cells (Fig. 4b; compare the 80S peak to the peak at the top of the gradient). Although the 60S subunit deficiency of *ubi1* cells is moderate, it is presumably sufficient to account for the mild

growth defect of these cells (Fig. 1c), which still express a protein identical to *UBI1* from the *UBI2* gene.

The partial suppression of the *ubi3* growth defect by the *ubi1* and *ubi2* mutations (Fig. 1c) may result from their lowering of 60S subunit levels and thus partially alleviating the relative deficiency of 40S subunits, perhaps without significantly affecting the absolute levels of 40S subunits. Although the coupling of 40S and 60S subunits within the ribosome only occurs after the 40S subunit associates with mRNA, the two subunits have significant residual affinity for each other in the absence of mRNA³². As a result, the abnormally high level of free 60S subunits in the *ubi3* mutant (Fig. 4a) may lead to partial sequestration of 40S subunits, thereby further decreasing the concentration of free 40S subunits available for translation in *ubi3* cells. This effect, analogous to the 'squelching' effect described for transcriptional regulators³³, should become weaker after a decrease in the level of 60S subunits, possibly accounting for the suppressive effect of the *ubi1* and *ubi2* mutations in the *ubi3* genetic background (Figs 1c and 3). As the growth of the *ubi3* mutant is cold-sensitive (relative to wild-type), whereas the ratio of 18S rRNA to 25S rRNA in *ubi3* cells does not change at low temperatures (data not shown), the hypothetical sequestration effect may decrease at higher temperatures. Alternatively, those few 40S ribosomal subunits which do form in the absence of the *UBI3* tail may be intrinsically

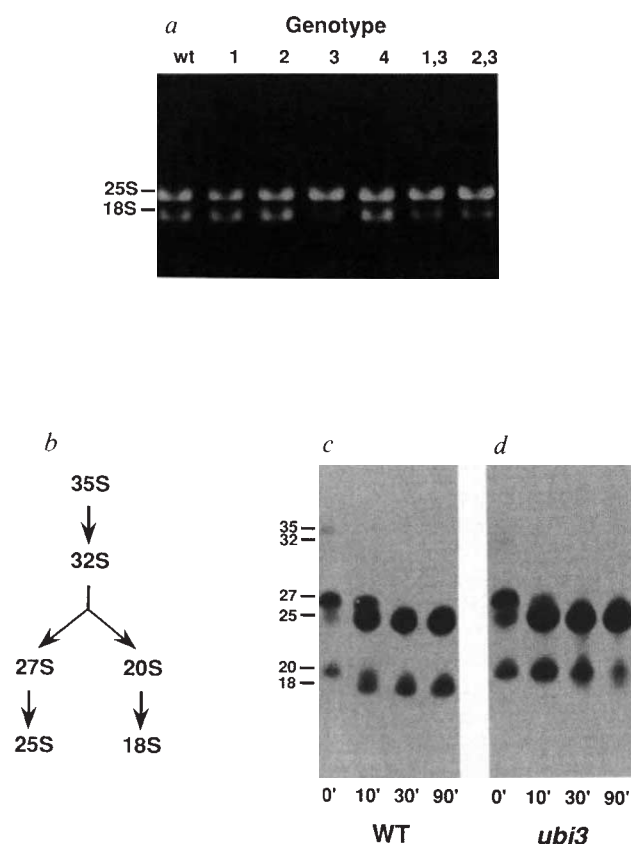


FIG. 3 Defective pre-rRNA processing and reduced levels of 18S rRNA in the *ubi3* mutant. *a*, RNA was prepared from mutant and wild-type cells grown in YPD, electrophoresed on a 1% agarose gel in the presence of ethidium bromide and photographed under ultraviolet light. Lanes are numbered according to the ubiquitin gene or genes deleted in the corresponding strain. *b*, Schematic representation of the pathway of pre-rRNA processing in *S. cerevisiae*. 35S RNA is the primary transcript. *c*, *d*, Pulse-chase analysis of pre-rRNA processing in wild-type and *ubi3* deletion strains. rRNA precursors in exponentially growing yeast cells were labelled (in minimal medium supplemented with auxotrophic nutrients⁴⁶) with [methyl-³H]methionine (Amersham, 80 Ci mmol⁻¹) for 6 min at 23 °C. ([methyl-³H] methionine donates methyl groups to newly synthesized pre-rRNA²⁹ via the general methyl donor *S*-adenosylmethionine.) Unlabelled methionine (1 mg ml⁻¹) was added to initiate the chase incubation, and RNA was prepared from the cells at the times indicated. RNA was electrophoresed on a 1% agarose gel and visualized by fluorography. Molecular weights are shown on the left, in thousands. **METHODS.** RNA was prepared from exponentially growing cells as described previously⁵², except that 0.2% diethylpyrocabonate was added to the lysis buffer and SDS was omitted. RNA (10 µg per lane) was electrophoresed on formaldehyde-containing gels as described previously⁴⁷, except that formaldehyde was used at 0.2 M, and iodoacetamide (10 mM) and ethidium bromide (0.5 µg ml⁻¹) were also added to the gel (B. Seed, personal communication). All strains analysed were *MATα*.

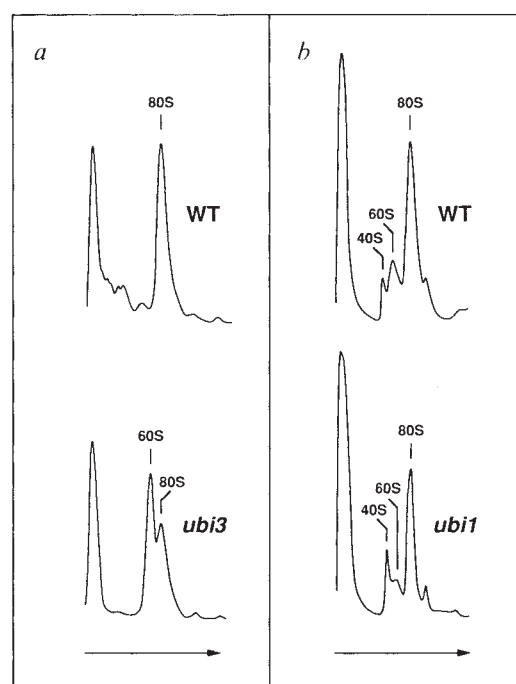


FIG. 4 Aberrant stoichiometries of ribosomal subunits in *ubi1* and *ubi3* mutants. Sedimentation analysis of subunits present in cytoplasmic extracts from spheroplasts of either (*a*) wild-type and *ubi3* strains, or (*b*) wild-type and *ubi1* strains. Direction of sedimentation is indicated by an arrow. **METHODS.** Spheroplasts were prepared from cells grown in YPD by digestion with zymolyase as described previously³¹. After allowing the spheroplasts to recover³¹, sodium azide was added, and the culture was brought to 8 °C for 5 min to induce polyribosome 'run off'³¹. Spheroplasts were then lysed as described³¹, except that heparin (0.2 mg ml⁻¹) was added to the lysis buffer⁵³. Extracts were centrifuged at 20,000 *g* for 15 min. In *b*, the ionic strength of the extract was then reduced by diluting it threefold into gradient buffer. Supernatants were loaded onto 15–30% linear sucrose gradients prepared to 50 mM Tris-HCl (pH 7.4), 50 mM KCl, 5 mM MgCl₂, 2 mM dithiothreitol and centrifuged at either 39,000 r.p.m. for 3 h (*a*) or 35,000 r.p.m. for 5 h (*b*) in an SW41 rotor (Beckman) at 5 °C. Gradients were fractionated by upward displacement in an ISCO gradient fractionator equipped with a UV monitor. Peak assignments are based on electrophoretic analysis of RNA prepared from appropriate gradient fractions. All strains analysed were *MATα*.

cold-sensitive, presumably because they lack the UBI3 tail (see below).

The tails are ribosomal proteins

The above results, which indicate that the UBI3 tail participates in ribosome biogenesis, suggested that this tail was either a component of the pre-rRNA processing apparatus, which seems to include a variety of small nuclear ribonucleoprotein particles³⁴, or a ribosomal protein whose incorporation into the nascent ribosome facilitates recognition of the 20S pre-rRNA by specific processing factors. To distinguish between these possibilities, we added an immunological marker to the C-terminus of the UBI3 protein by fusing the *UBI3* gene with a sequence encoding a 12-residue epitope from the human *c-myc* gene product³⁵. The resulting *UBI3-myc* gene fully complemented the growth defect of the chromosomal *ubi3* deletion, indicating that the C-terminal Myc peptide did not obstruct the activity of the UBI3 protein. Whole-cell extracts of wild-type (*UBI3*) cells and of *UBI3-myc* cells carrying a chromosomal deletion of *ubi3* were fractionated by SDS-PAGE, transferred onto filters and probed with a monoclonal antibody to the Myc peptide. A single band was observed in immunoblots of *UBI3-myc* cells, whereas extracts from wild-type cells were essentially free of antibody-binding species (Fig. 5a, lane W). The immunoreactive protein is apparently the deubiquitinated form of UBI3-Myc, as it is not recognized by antibodies to ubiquitin, and its apparent molecular weight is considerably less than that of UBI3-Myc produced in *Escherichia coli* (Fig. 5a). This, with the data of Fig. 2, suggests that the UBI3 hybrid protein is efficiently processed (deubiquitinated) *in vivo*. Rapid *in vivo* deubiquitination in eukaryotic cells (but not in bacteria) is also characteristic of engineered ubiquitin fusions such as ubiquitin- β -galactosidase³⁶.

Extracts from wild-type and *UBI3-myc* cells were analysed on sucrose gradients under optimal conditions for the resolution of polyribosomes (Fig. 5a). The proteins of the monosome and polyribosome peaks were fractionated by SDS-PAGE, transferred onto filters and probed with the anti-Myc antibody. The Myc-tagged UBI3 tail protein was found primarily within rapidly sedimenting structures, and co-sedimented with both 80S monosomes and the broadly distributed polyribosomes (Fig. 5a). A second extract was prepared from *UBI3-myc* cells under conditions that dissociate polyribosomes, and centrifuged in the presence of a high ionic strength buffer, which dissociates monosomes into free subunits. Under these conditions the UBI3 tail co-sedimented with the 40S ribosomal subunit (Fig. 5b). Taken together, these data indicate that the UBI3 tail is a component of the small subunit of the ribosome. The UBI3 tail does not seem to be a processing or assembly factor that is transiently associated with nascent 40S subunits because it is also present within translationally active ribosomes. The association of the UBI3 tail with the 40S subunit persists under high salt conditions (0.5 M KCl), suggesting that the UBI3 tail is tightly associated with mature ribosomes.

To localize the UBI1 tail protein, we again used the Myc tagging method. A *UBI1 tail-myc* gene in which the Myc sequence was positioned upstream of the tail-coding sequence in place of the ubiquitin-coding sequence, rescued the viability of *ubi1 ubi2* mutants and restored their growth rates to near wild-type levels. Extracts from *UBI1 tail-myc* cells were fractionated on sucrose gradients optimized for the analysis of either polyribosomes or ribosomal subunits, as described above for *UBI3-myc*. On the polyribosome gradient, the UBI1 tail co-sedimented with 80S monosomes and polyribosomes (Fig. 5c), whereas on the ribosomal subunit gradient the UBI1 tail co-sedimented with the 60S subunit (Fig. 5d). Thus, like the UBI3 tail, the UBI1 tail is tightly associated with mature, translationally active ribosomes. We conclude that the UBI3 tail, and the identical UBI1 and UBI2 tails, are ribosomal proteins which reside in the small and large subunits, respectively. The main

structural motifs of the UBI1-UBI3 proteins, their putative zinc fingers²⁴, are presumably involved in binding rRNA.

Although both UBI1 tail-Myc and UBI3-Myc were found in ribosomes, their localization to distinct ribosomal subunits precludes the possibility that these associations were mediated artefactually by the Myc peptide. Moreover, the human homologue of the UBI3 tail protein has independently been localized to the small ribosomal subunit³⁷. Our data do not exclude the possibility that a fraction of the UBI1 and UBI2 tails are present within yeast ribosomes in an unprocessed, ubiquitin-containing form. The presence of significant levels of free ubiquitin in *ubi3 ubi4* mutants (Fig. 1c) does indicate, however, that the UBI1 and UBI2 gene products can be deubiquitinated; furthermore, the full complementing activity of the free UBI1 tail indicates that ribosomes containing the free UBI1 tail are functionally indistinguishable from wild-type.

Examination of the previously determined²⁴ nucleotide sequences upstream of the *UBI1*, *UBI2* and *UBI3* genes revealed near-matches to the consensus sequence³⁰ of upstream activation sites (UAS) within genes for ribosomal proteins (Fig. 1b). (Such UAS^{TP8} sites have also been found within a few genes encoding non-ribosomal proteins³⁰.) The *UBI1*, *UBI2* and *UBI3* promoter regions also contain T-rich stretches which represent a second sequence motif characteristic of promoters of ribosomal protein genes³⁰ (Fig. 1b). In *UBI1-UBI3*, the positioning of these UAS and T-rich sequences with respect to each other and to sites of translational initiation, is similar to that of known ribosomal protein genes and consistent with the proposed function of these sequences in transcriptional regulation. With the exception of genes for ribosomal proteins, few yeast genes contain introns; their presence within the *UBI1* and *UBI2* genes (Fig. 1a) further indicates that the sequence features of *UBI1-UBI3* are typical of ribosomal protein genes.

Ubiquitin and ribosome biogenesis

The tandem arrangement of ubiquitin and specific ribosomal proteins (tails) within proteolytically processed precursors has been conserved throughout the evolution of eukaryotes^{7,25}. Nevertheless, expression of the tails in a ubiquitin-free form can fully complement chromosomal mutations in which the corresponding ubiquitin-tail gene is entirely deleted (see above). To test further for a function of the cotranslational synthesis of ubiquitin and the tail proteins, we integrated the free tail-expressing derivative of the *ubi3* gene (*ubi3 Δ ub*) into its natural chromosomal locus. The complete deletion mutant, *ubi3*, which grows slowly (Fig. 1c), was transformed with linear DNA fragments carrying either the ubiquitin deletion mutation *ubi3 Δ ub* or the wild-type *UBI3* gene and growth revertants were selected. Growth revertants resulting from transformation with *ubi3 Δ ub* DNA were predominantly of two classes: those growing at wild-type rates (1.6-h doubling time) and those in which the *ubi3* growth defect had been only partially reverted (3-h doubling time, compared with the 6.8-h doubling time of the *ubi3* mutant). Southern hybridization analysis (data not shown) indicated that in all seven of the revertants with wild-type growth rates, the *ubi3* mutation had been replaced by multiple tandem insertions of *ubi3 Δ ub* DNA, whereas in all 12 partial growth revertants analysed, *ubi3* had been replaced with a single, properly integrated *ubi3 Δ ub* sequence (see Fig. 6a for a comparison of growth rates; we refer to the integrated, single-copy *ubi3 Δ ub* allele as *ubi3 Δ ub-1*, and to the integrated multiple-copy allele taken for analysis as *ubi3 Δ ub-2*). In contrast, single-copy integration of wild-type (*UBI3*) DNA resulted in complete reversion of the *ubi3* growth defect (Fig. 6a). These results, which strongly suggested that the *ubi3 Δ ub* allele can fully complement *ubi3* only when present in multiple copies, were confirmed by showing that in a *ubi3* genetic background, the copy number of CEN plasmids carrying the *ubi3 Δ ub* allele was several times higher than that of CEN plasmids carrying the *UBI3* allele, whereas in a wild-type genetic background such

plasmids had equivalent copy numbers (data not shown). Thus, regardless of whether the free tail-expressing (*ubi3Δub*) allele is carried on a *CEN* plasmid or integrated into the chromosome, full complementation of the *ubi3* phenotype requires approximately three to fourfold amplification of the *ubi3Δub* copy number.

To determine the stoichiometry of 18S to 25S rRNA in the *ubi3Δub-1* mutant (which contained the integrated, single-copy *ubi3Δub* allele), RNA from this and control strains was electrophoresed on an agarose gel and visualized by ethidium staining (Fig. 6b). A deficiency of 18S rRNA was clearly visible in the *ubi3Δub-1* mutant, although it was not as pronounced as in the *ubi3* mutant (Fig. 6b). A wild-type (1:1) stoichiometry of 18S rRNA to 25S rRNA was seen in the *ubi3Δub-2* mutant, which contained multiple copies of *ubi3Δub* (Fig. 6b). Thus, deletion of the ubiquitin-coding element of *UBI3* leads to a deficiency of 18S rRNA which can be suppressed by amplifying the copy number of the free tail-expressing allele of the *UBI3* gene. This indicates that the ubiquitin-coding portion of the *UBI3* gene, although not strictly required for the processing of

20S pre-rRNA, facilitates this processing reaction, apparently by increasing the efficiency of incorporation of the *UBI3* tail into the nascent 40S ribosomal subunit.

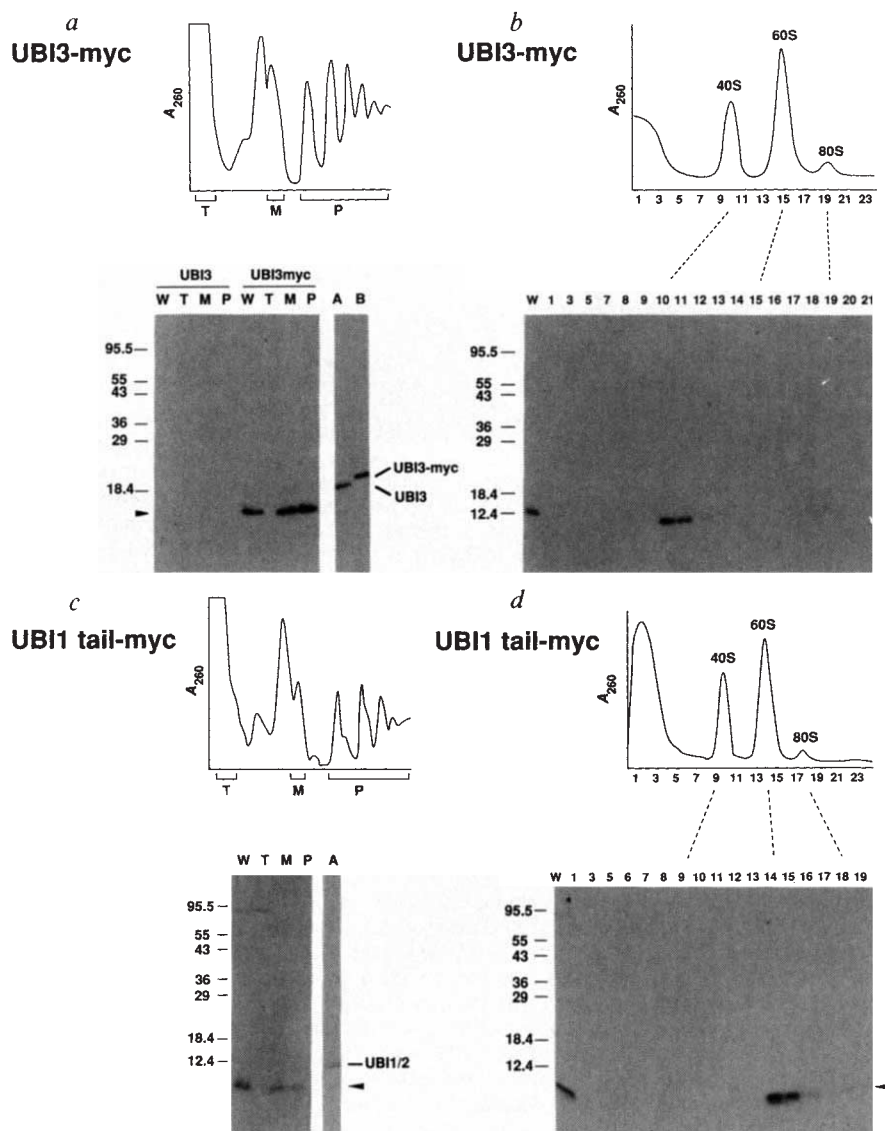
To test whether the effect of deleting the ubiquitin coding sequence of *UBI3* on tail function was achieved through a post-translational mechanism, *UBI3* mRNA levels were determined in *ubi3Δub-1* and control strains. The northern hybridization analysis of Fig. 6c showed that the effect of the *ubi3Δub-1* mutation on *UBI3* tail function is not achieved by decreasing the level of *UBI3* mRNA; instead, the *ubi3Δub-1* mRNA is considerably more abundant than that of *UBI3*. Thus, the primary functional defect of the *ubi3Δub-1* mutant lies in either a cotranslational or post-translational process. We conclude that the effective levels of newly synthesized *UBI3* tail protein, as reflected by its capacity to accelerate the processing of 20S pre-rRNA (Figs 3, 6), are significantly enhanced by its transient covalent association with ubiquitin.

Discussion

Ubiquitin has previously been shown to function as a post-

FIG. 5 Immunoblot analysis of *S. cerevisiae* extracts fractionated in sucrose gradients to resolve either polyribosomes (a, c) or dissociated 40S and 60S subunits (b, d). In panels a and c, fractions were pooled as shown: W, whole-cell extract; T, top of gradient; M, monosomes; P, polyribosomes. a, Polyribosomal fractions from *ubi3* cells carrying either *UBI3* (lanes 1–4) or *UBI3-myc* (lanes 5–8) on a *CEN* plasmid, probed with anti-Myc antibody. Lanes A and B show the unprocessed (non-deubiquitinated) *UBI3* and *UBI3-myc* produced in *E. coli*, and probed with anti-ubiquitin antibody²¹. b, Ribosomal subunit fractions from *ubi3* cells carrying *UBI3-myc* on a *CEN* plasmid, probed with anti-Myc antibody. c, Polyribosomal fractions from *ubi1 ubi2* cells carrying *UBI1 tail-myc* on a 2 μ plasmid, probed with anti-Myc antibody. Lane A shows unprocessed *UBI1* produced in *E. coli*, and probed with anti-ubiquitin antibody. d, Ribosomal subunit fractions from *ubi1 ubi2* cells carrying *UBI1 tail-myc* on a 2 μ plasmid, probed with anti-Myc antibody. In c, the monosome and polyribosome peaks each have a shoulder of more rapidly sedimenting material. Similar results have been obtained with unrelated mutants deficient in large ribosomal subunits⁵⁴, as is the *ubi1* strain (Fig. 4). The shoulder seems to represent mRNAs which have bound the 40S ribosomal subunit to form a preinitiation complex, but have not proceeded to bind a 60S subunit because the level of free 60S subunits is low⁵⁴. Molecular weights are shown in thousands.

METHODS. To construct pUb54, the 1.3-kb *HincII* fragment from *UBI3* was ligated to the 0.7-kb *HincII*–*SphI* fragment of the plasmid AHSM (ref. 55) and inserted into YCp50 (ref. 50). The resulting plasmid encodes *UBI3-myc*, in which the last two amino acids of *UBI3* are replaced by the 12-residue Myc tag. To construct pUb62, the ~30-bp *Bam*HI–*Hind*III fragment of pUb92, encoding the first seven residues of the *UBI1* tail, was replaced by a synthetic oligonucleotide encoding the 12-residue Myc tag followed by the first seven residues of the *UBI1* tail. (A C-terminal fusion of Myc to *UBI2* did not complement the *ubi1 ubi2* double mutation; data not shown). The resulting plasmid was cut with *Eco*RI and *Eco*RV and the 2.3-kb fragment containing *UBI1 tail-myc* was cloned into the *Sma*I site of YEp351 (ref. 49), yielding pUb62. For the preparation of polyribosomes, cells were grown to an A_{600} ~0.5. Extracts were prepared³¹ and analysed⁵³ as described. Gradients were centrifuged in an SW41 rotor at 15,000 r.p.m. for 16 h. Proteins were precipitated with 25% trichloroacetic acid, the pellets washed, dissolved and electrophoresed and electrophoretically transferred as in Fig. 1. Filters were probed with the Myc1-9E10 antibody³⁵ followed by ¹²⁵I-labelled sheep



anti-mouse IgG, using standard techniques⁸. Preparation of extracts for the analysis of ribosomal subunits was as in the Fig. 4 legend, except that extracts were brought to 0.5 M KCl before loading onto 15–35% sucrose gradients prepared as in the Fig. 4 legend but with the addition of 0.5 M KCl. Gradients were centrifuged in an SW41 rotor at 21,000 r.p.m. for 17 h at 10 °C.

translational modifying group which signals the degradation of acceptor proteins. Here, we describe a second mechanism of ubiquitin function in which ubiquitin is cotranslationally joined to an acceptor protein and serves to promote the assembly of the acceptor protein into a specific cellular structure. The role we have described for ubiquitin in the biogenesis of ribosomes in yeast probably applies to eukaryotes in general, as close homologues of the *UBI1-UBI3* genes exist in animals, plants

and lower eukaryotes^{7,25}. Moreover, sequences homologous to the tail components of *UBI1-UBI3* have invariably been found fused to ubiquitin.

UBI1-UBI3 form a subset of a class of genes in which an element encoding either ubiquitin or an amino-acid sequence similar to that of ubiquitin is fused to a downstream coding sequence. Although no systematic effort to identify such genes has yet been reported, four loci encoding C-terminally extended ubiquitin-like polypeptides have already been described. These are: first, the *An-1* gene, whose transcripts are segregated to the animal pole of *Xenopus laevis* embryos (D. Weeks and D. Melton, personal communication); second, a human gene that encodes an interferon-inducible protein with a relative molecular mass of about 15,000 (refs 38, 39); third, the constitutively expressed human gene *GdX* (ref. 40); and finally, the large open reading frame of the togavirus BVDV (bovine viral diarrhoea virus), which encodes a ubiquitin-like sequence embedded within the nonstructural viral polyprotein⁴¹. Because of their substantial divergence from the canonical, highly conserved ubiquitin sequence, the ubiquitin-like proteins cannot serve as precursors to ubiquitin, in contrast to the *UBI1-UBI3* proteins. Moreover, the processing proteases that remove ubiquitin from ubiquitin fusion proteins are expected to be virtually inactive upon the ubiquitin-like proteins, because the bond that would be cleaved is flanked by amino-acid substitutions that block (or are expected to block) cleavage of engineered ubiquitin fusion proteins^{36,42}. We suggest that the function of ubiquitin-like protein domains may be to enhance the function of linked non-ubiquitin domains in a way similar to that seen in the *UBI3* protein.

It has recently been shown that certain proteins, known as 'molecular chaperones', associate transiently (and non-covalently) with newly synthesized polypeptide chains and so facilitate their incorporation into oligomeric structures⁴³⁻⁴⁵. By definition, molecular chaperones do not form part of the final oligomeric structure, nor do they necessarily possess steric information specifying the structure's assembly⁴³. By these criteria, it seems that the ubiquitin moiety of the *UBI3* protein acts as a molecular chaperone of the *UBI3* tail in facilitating its incorporation into ribosomes. The covalent, cotranslational association of a chaperone protein with its ligand has several potential advantages: the N-terminal chaperone is bound to and can protect its ligand from the moment the ligand is synthesized; the chaperone can transiently block, or shelter, the N-terminus of its ligand (which may serve as a signal for degradation in some proteins³⁶); and finally, the cotranslational chaperone can bind its ligand without needing a specific recognition site in the ligand.

It is probable that the ubiquitin tails, like other ribosomal proteins³⁰, are extremely short-lived if they are not assembled within ribosomes. Thus, the transient attachment of ubiquitin to the N-termini of the *UBI1-UBI3* tails could increase the efficiency of their incorporation into nascent ribosomes by a transient metabolic stabilization of the newly formed tails. Although ubiquitination of a protein can signal its degradation, this activity seems to require multi-ubiquitin chains²³. Rather than directly protecting the *UBI1-UBI3* tails against degradation, ubiquitination could increase the rate of the tails' transport to, or assembly within, nascent ribosomes, thus decreasing the proportion of newly synthesized tails that are degraded in transit to the ribosome. Experiments to distinguish between these possibilities are in progress. □

Received 12 December 1988; accepted 9 February 1989.

1. Hershko, A., Ciechanover, A., Heller, H., Haas, A. L. & Rose, I. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1783-1786 (1980).
2. Finley, D., Ciechanover, A. & Varshavsky, A. *Cell* **37**, 43-55 (1984).
3. Ciechanover, A., Finley, D. & Varshavsky, A. *Cell* **37**, 57-66 (1984).
4. Jentsch, S., McGrath, J. P. & Varshavsky, A. *Nature* **329**, 131-134 (1987).
5. Goebel, M. G. *et al. Science* **241**, 1331-1335 (1988).
6. Kulka, R. G. *et al. J. Biol. Chem.* **263**, 15726-15731 (1988).
7. Schlesinger, M. J. & Bond, U. *Oxford Surveys on Eukaryotic genes* **4**, 77-91 (1987).

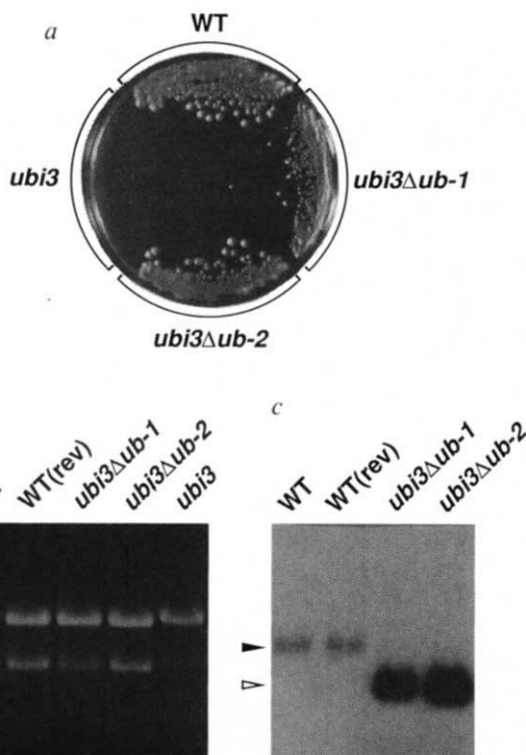


FIG. 6 Characterization of mutants bearing single or multiple copies of the *ubi3Δub* allele. *a*, Growth defect of the *ubi3Δub-1* mutant. Cultures were streaked onto YPD plates and incubated for 3 days at 30 °C. Quadrant designations are as follows: WT, wild-type; *ubi3Δub-1*, a growth revertant obtained through integrative transformation of a *ubi3* strain with a DNA fragment carrying the *ubi3Δub* deletion mutation (single-copy integration); *ubi3Δub-2*, as *ubi3Δub-1*, except that 3-4 copies of the transforming DNA fragment integrated tandemly at the *UBI3* locus; *ubi3*, which forms sizeable colonies only after 1 week. *b*, 18S rRNA deficiency in the *ubi3Δub-1* mutant. RNA was resolved on a 1% agarose gel in the presence of ethidium bromide, and photographed under UV light. Designations are as in *a*, except for lane WT (rev), which contained RNA from a phenotypically wild-type revertant obtained through integrative transformation of a *ubi3* strain with a DNA fragment carrying the wild-type *UBI3* gene. *c*, Northern hybridization analysis of *UBI3* RNA levels in mutant strains. Closed arrow, *UBI3* RNA; open arrow, *ubi3Δub* RNA.

METHODS. Growth revertants were selected on YPD plates at 23 °C, a temperature at which the (cold-sensitive) *ubi3* mutants form colonies only after more than two weeks. Transformants were screened for loss of the *ubi3*-linked *HIS3* marker before Southern hybridization analysis. *b* and *c*, RNA was prepared from cells grown in YPD, and electrophoresed as in Fig. 3, except that in *c* a 1.5% agarose gel was used, then transferred onto a Gene Screen filter (New England Nuclear), and UV-irradiated as described⁸. Hybridization was carried out at 37 °C in a solution containing 40% formamide, 7% SDS, and 5 × SSPE. The filter was washed at 62 °C in 1% SDS, 1 × SSPE. The *UBI3* hybridization probe was the insert from pUB36 (see the Fig. 2 legend). The probe was radiolabelled as described⁸. In *c*, 10 μg of RNA was loaded onto each lane; *b*, loads were slightly adjusted to equalize the intensities of the 25S rRNA bands to facilitate comparison of 25S rRNA:18S rRNA ratios between the lanes. To increase the sensitivity of detection, formaldehyde was omitted from the gel in *b*. In *c*, only one *UBI3* mRNA (of ~0.7 kb) was detectable in wild-type cells. This and other data, including an analysis of transcripts in the *ubi3* mutant (data not shown), indicate that the 1.8 kb and 1.0 kb mRNAs previously considered to be additional *UBI3* transcripts²⁴ are not in fact derived from the *UBI3* locus.

8. Finley, D., Özkaynak, E. & Varshavsky, A. *Cell* **48**, 1035–1046 (1987).
9. Tanaka, K., Matsumoto, K. & Toh-e, A. *EMBO J.* **7**, 495–502 (1988).
10. Treger, J. M., Heichman, K. A. & McEntee, K. *Molec. cell. Biol.* **8**, 1132–1136 (1988).
11. Goff, S. A., Voellmy, R. & Goldberg, A. in *Ubiquitin* (ed. Rechsteiner, M.) 207–238 (Plenum, New York, 1988).
12. Herskko, A. *J. biol. Chem.* **263**, 15237–15240 (1988).
13. Varshavsky, A., Bachmair, A., Finley, D., Gonda, D. & Wüning, I. in *Ubiquitin* (ed. Rechsteiner, M.) 287–324 (Plenum, New York, 1988).
14. Finley, D. & Varshavsky, A. *Trends Biochem. Sci.* **10**, 343–346 (1985).
15. Pickart, C. M. in *Ubiquitin* (ed. Rechsteiner, M.) 77–100 (Plenum, New York, 1988).
16. Bonner, W. M., Hatch, C. L. & Wu, R. S. in *Ubiquitin* (ed. Rechsteiner, M.) 157–172 (Plenum, New York, 1988).
17. Ball, E. *et al. Cell* **51**, 221–228 (1987).
18. Siegelman, M. *et al. Science* **231**, 823–829 (1986).
19. Yarden, T. *et al. Nature* **323**, 226–232 (1986).
20. Leung, D. W. *et al. Nature* **330**, 537–543 (1987).
21. Perry, G., Friedman, R., Shaw, G. & Chau, V. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3033–3036 (1987).
22. Manetto, V. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4501–4505 (1988).
23. Chau, V. *et al. Science* (in the press).
24. Özkaynak, E., Finley, D., Solomon, M. J. & Varshavsky, A. *EMBO J.* **6**, 1429–1439 (1987).
25. Finley, D. *et al. in Ubiquitin* (ed. Rechsteiner, M.) 39–75 (Plenum, New York, 1988).
26. Ecker, D. J., Khan, M. I., Marsh, J., Butt, T. R. & Crooke, S. T. *J. biol. Chem.* **262**, 3524–3527 (1987).
27. Bitter, G. A. & Egan, K. M. *Gene* **69**, 193–207 (1988).
28. Lund, P. K. *et al. J. biol. Chem.* **260**, 7609–7613 (1985).
29. Warner, J. R. in *The Molecular Biology of the Yeast Saccharomyces cerevisiae: Metabolism and gene Expression* (eds Strathern, J., Jones, E. & Broach, J.) 529–560 (Cold Spring Harbor Laboratory, New York, 1981).
30. Planta, R. J. & Raué, H. A. *Trends Genet.* **4**, 64–68 (1988).
31. Carter, C. J., Cannon, M. & Jiménez, A. *Eur. J. Biochem.* **107**, 173–183 (1980).
32. van Holde, K. E. & Hill, W. E. in *Ribosomes* (eds Nomura, M., Tissieres, A. & Lengyel, P.) 53–91 (Cold Spring Harbor Laboratory, New York, 1974).
33. Gill, G. & Plashne, M. *Nature* **334**, 721–724 (1988).
34. Tollervey, D. *EMBO J.* **6**, 4169–4175 (1987).
35. Munro, S. & Pelham, H. R. B. *Cell* **46**, 291–300 (1986).
36. Bachmair, A., Finley, D. & Varshavsky, A. *Science* **234**, 179–186 (1986).
37. Redman, K. & Rechsteiner, M. *Nature* (in the press).
38. Blomstrom, D. C., Fahey, D., Kutney, R., Korant, B. D. & Knight, E. Jr. *J. Biol. Chem.* **261**, 8811–8816 (1986).
39. Haas, A. L., Ahrens, P., Bright, P. M. & Ankel, H. *J. biol. Chem.* **262**, 11315–11323 (1987).
40. Toniolo, D., Persico, M. & Alcalay, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 851–855 (1988).
41. Clarke, L. E. *et al. Nature* **337**, 709–716 (1989).
42. Butt, T. R., Khan, M. I., Marsh, J., Ecker, D. J. & Crooke, S. T. *J. biol. Chem.* **263**, 16364–16371 (1988).
43. Ellis, J. *Nature* **328**, 378–379 (1987).
44. Hemmingsen, S. M. *et al. Nature* **333**, 330–334 (1988).
45. Laemmli, U. K. & Favre, M. *J. molec. Biol.* **80**, 575–599 (1973).
46. Sherman, F., Fink, G. R. & Hicks, J. B. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, New York, 1986).
47. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
48. Vieira, J. & Messing, J. *Gene* **19**, 259–268 (1982).
49. Hill, J. E., Myers, A. M., Koerner, T. J. & Tzagoloff, A. *Yeast* **2**, 163–167 (1986).
50. Parent, S. A., Fenimore, C. M. & Bostian, K. A. *Yeast* **1**, 83–138 (1985).
51. Guarente, L., Yocum, R. R. & Gifford, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7410–7414 (1982).
52. Sprague, G. F. Jr., Jensen, R. & Hershkowitz, I. *Cell* **32**, 409–415 (1983).
53. Warner, J. R., Mitra, G., Schwindinger, W. F., Studeny, M. & Fried, H. M. *Molec. cell. Biol.* **5**, 1512–1521 (1985).
54. Rotenberg, M. O., Moritz, M. & Woolford, J. L. Jr. *Genes Dev.* **2**, 160–172 (1988).
55. Munro, S. *thesis*, Univ. Cambridge (1987).

ACKNOWLEDGEMENTS We thank Mary Jalenak for assistance in characterizing mutants, John McGrath for carrying out sequence searches, Joan Park and Andreas Bachmair for helpful discussions, Vincent Chau for the anti-ubiquitin antibody, Tauseef Butt and David Ecker for the plasmid YEp46, Sean Munro for advice on peptide tagging, Brian Seed for advice on RNA electrophoresis, members of this laboratory for comments on the manuscript, Daniel Weeks and Douglas Melton for permission to cite unpublished data, and Barbara Doran for secretarial assistance. This work was supported by grants to A.V. from the NIH. B.B. was supported by a predoctoral fellowship from the NSF.

LETTERS TO NATURE

An alternative binary model for SN1987A

Ph. Podsiadlowski & P. C. Joss

Department of Physics, Center for Space Research, and Center for Theoretical Physics, Massachusetts Institute of Technology, Room 6-203, Cambridge, Massachusetts 02139, USA

FABIAN *et al.*¹ and Joss *et al.*² have proposed that the progenitor of supernova 1987A was a relatively faint binary companion of the blue supergiant Sk – 69°202 (ref. 3). Here we propose another class of models for SN1987A, wherein Sk – 69°202 was the progenitor, but where the presupernova evolution of this star was strongly influenced by the accretion of matter from a companion star which was initially the more massive star. Binary stellar evolution calculations show that a model of this type can naturally explain why the apparent progenitor was a blue supergiant rather than a red one. The system becomes unbound after the supernova event, and the presence of a normal or neutron-star companion will probably be revealed a few years after the explosion. A model of this type can explain all the major observational features of SN1987A and also provides plausible explanations for many of its anomalies.

SN1987A has enriched many fields of astrophysics as well as some related disciplines. Surprisingly, however, the greatest challenge it has posed is to stellar evolution theory: why was the apparent progenitor (Sk – 69°202; ref. 3) a blue supergiant rather than a red supergiant, the generally preferred precursor of type II supernovae^{4,5}? Numerous single-star models for the progenitor have been advanced to explain this unexpected behaviour, but all of them face significant objections (for a review of single-star models see, for example, ref. 6 and references therein).

The resolution of this puzzle is of great importance. An understanding of the presupernova evolution would allow us to decide whether SN1987A was a unique event or was representative of a whole class of supernovae which are too faint to be easily detected outside the Local Group. Moreover, if this supernova is telling us something about the evolution of massive stars, it may turn out to be a Rosetta stone against which the theory of stellar evolution can be calibrated.

Most stars occur in binary or multiple systems⁷. If we allow that the supernova progenitor may have been a member of a binary, several alternative models present themselves. In one class of models, the star that exploded was not Sk – 69°202, but a previously undetected companion star^{1,2}. These models predict that Sk – 69°202 will reappear when the photosphere of the supernova has receded sufficiently. In this letter we propose a different type of model, in which the progenitor was a blue rather than a red supergiant because of its past interaction with its companion. Our calculations also shed new light on the features that single-star models must have to account for the observed properties of SN1987A.

Whether a star ends its life as a red or a blue supergiant depends on some rather subtle aspects of the physics of the stellar envelope and is not entirely amenable to heuristic explanation. However, many of the essential features of the envelope physics can be described in terms of how much luminosity can be carried in the envelope by radiative diffusion (refs 8, 9; but see also refs 10 and 11). If the luminosity of a star during the course of its evolution reaches a critical value, L_{crit} , a thermal instability triggers a runaway expansion of the envelope (ref. 8) and drives the star redward toward the Hayashi track. L_{crit} is a sensitive function of the radiative opacities in the envelope and increases with decreasing opacities, a result that accounts for the requirement of very low metallicities in most single-star models for SN1987A. L_{crit} is also very sensitive to the fractional core mass $q_c \equiv M_c/M$, where M_c is the mass of the hydrogen-exhausted core and M is the total mass of the star. It is well known from the theory of the evolution of massive stars that a reduction in envelope mass (for example, by mass loss through a stellar wind) and a consequent increase in q_c reduces L_{crit} and favours redward motion in the Hertzsprung–Russell diagram¹². Conversely, a decrease in q_c (for example, by accretion of matter from a close-binary companion) increases L_{crit} , and the transition to the red-supergiant stage may thereby be prevented (or reversed if it has already occurred).

To determine whether this mechanism may explain the seemingly anomalous colour of Sk – 69°202, we have performed a series of stellar evolution calculations in which such accretion is allowed to occur. We used a standard Henyey-type code¹³, and we assumed an initial metallicity of $Z = 0.02$ and that convection was governed by the Schwarzschild instability

criterion, with a convective mixing length equal to 1.5 pressure scale heights, and with the effects of convective overshooting and semiconvection included according to the prescription given by Iben¹⁴. We followed the evolution of three stellar models with initial masses of 15, 17 and 20 $M_{\odot}$ from the zero-age main sequence to the onset of core carbon burning. After the termination of the core hydrogen-burning phase, we accreted 5 $M_{\odot}$ onto the 15- $M_{\odot}$ model and 3 $M_{\odot}$ onto the 17- $M_{\odot}$ model. Thus, all three models had a final mass of 20 $M_{\odot}$; the masses of the hydrogen-exhausted cores were 4.5, 5.3 and 6.4 $M_{\odot}$ respectively. We assume that the time that has elapsed since the end of the mass-transfer episode is long compared with the Kelvin time, $\tau_K \approx 10^4$ yr, of the progenitor, so that the appearance of the progenitor just before the supernova event should be essentially independent of the details of mass transfer.

The results of our calculations are illustrated in Fig. 1. Models with final values of q_c larger than a certain critical value q_c^0 are red supergiants (on the Hayashi track in the H-R diagram) just before becoming supernovae, whereas models with $q_c < q_c^0$ end their lives as blue supergiants. (For our chosen values of Z and final stellar mass, we find that $q_c^0 \approx 0.28$, but this value is sensitive to the parameters of the model.) The effective temperature of the star just before the supernova event increases monotonically (that is, the pre-supernova is increasingly blue) with increasing accreted mass.

Some of our detailed numerical results are sensitive to the assumed parameters of our models, but our qualitative con-

clusions regarding the viability of our binary models for SN1987A are insensitive to the assumed metallicity, the treatment of convection, the details of the mass-transfer process, and the amount of mass that may be lost by the supernova progenitor in a stellar wind following the mass-transfer phase. For our basic model to work, we require only that at least several solar masses more be accreted by the progenitor than are subsequently lost as a stellar wind. We also find that the dependence of the stellar luminosity on the core mass of the progenitor is much weaker than predicted by the core mass-luminosity relation for single massive stars in this evolutionary phase (see for example refs 15 and 16). We thus conclude that the luminosity does not accurately determine the core mass of a star that has accreted a substantial amount of matter after the termination of core hydrogen burning.

For the requisite binary evolution to occur, the supernova progenitor must have been originally the less massive star of a close binary system, because the original primary fills its critical lobe first and becomes the mass donor. From our evolutionary calculations, we also find that if mass transfer occurs before the supernova progenitor has completed core hydrogen burning, the progenitor will be rejuvenated as a main-sequence star and will subsequently mimic the standard evolution of a more massive single star; evolutionary scenarios are therefore also constrained by the requirement that the progenitor must complete core hydrogen burning before mass transfer begins. There are then two possible scenarios, involving case B (ref. 17) and case C (ref. 18) mass transfer, respectively, which have very different consequences.

In case B, the primary fills its critical lobe when it is on the first red-giant branch. As the supernova progenitor must have already left the main sequence before mass transfer begins, and as the main-sequence lifetime of a massive star is long compared with the interval from the termination of core hydrogen burning to the tip of the red-giant branch, then the original mass of the supernova progenitor must be very close to that of the primary. Using the results for our 15- $M_{\odot}$ model, we find that the original mass of the supernova progenitor must be within $\sim 0.3\%$ of the original mass of the primary. Binary systems composed of stars with the requisite masses and initial orbital separations (of the order of a few AU) will be rather rare, although a large fraction of all close binaries do in fact contain stars of approximately equal mass^{7,19}.

Because of the mass increase, the post-main-sequence evolution of the supernova progenitor is faster than the corresponding evolution of the original primary, and the progenitor will reach the supernova stage earlier than the primary (by as much as $\sim 10^5$ yr for our 15- $M_{\odot}$ model). In this case, Sk-69°202 would have had a stellar companion before the supernova event. However, as more than half the mass of the binary system should have been ejected in the supernova explosion, the system should not have remained bound, and the erstwhile companion should be revealed when or shortly after it crosses the photosphere of the supernova (unless the motion happens to be nearly directly away from the Earth). We estimate that this should occur about 1–3 yr after the explosion.

The constraint on the relative masses of the binary components for case C mass transfer is much weaker than in case B: the supernova progenitor must still be originally the less massive of the two, but its mass may differ from that of the original primary by as much as $\sim 10\%$. In this case, mass transfer occurs when the original primary reaches the asymptotic giant branch (after the completion of helium core-burning), whereas the supernova progenitor can be in any preceding post-main-sequence phase. The original primary then explodes as a supernova shortly after the mass-transfer phase and should leave a neutron-star remnant (this explosion occurred in the rather distant past and should not be confused with the observed supernova event). The binary system remains bound but acquires a small eccentricity, $e \approx 0.1$ –0.2. For up to some 10^6 yr after-

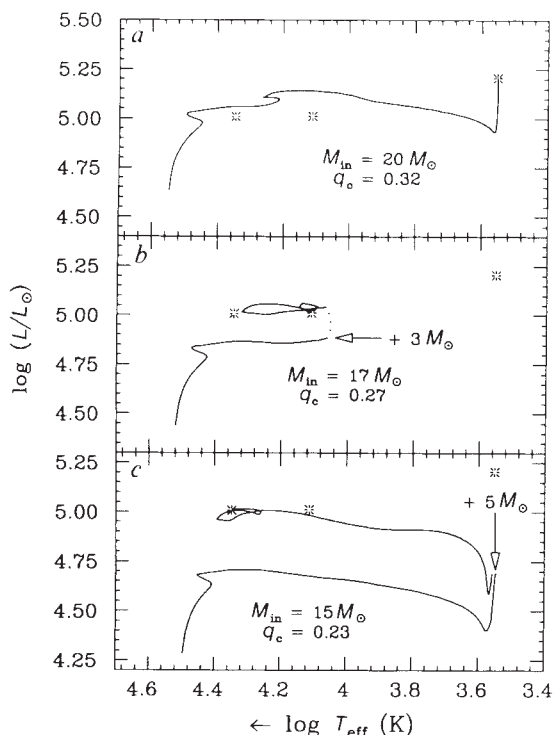


FIG. 1 Evolutionary tracks of the progenitor in the Hertzsprung-Russell diagram (stellar luminosity, L , versus effective temperature, T_{eff}) for several assumed amounts of accreted mass. *a*, Evolutionary track of a 20- $M_{\odot}$ star without accretion. *b*, Evolutionary track of a star with an initial mass of 17 $M_{\odot}$ which accretes 3 $M_{\odot}$ after the termination of core hydrogen burning (the dotted portion of the curve denotes the mass-transfer phase; the arrow indicates the onset of mass transfer). *c*, Evolutionary track for a star with an initial mass of 15 $M_{\odot}$ which accretes 5 $M_{\odot}$ after the core hydrogen-burning phase; here, the small break in the curve denotes the mass-transfer phase. In each case, M_{in} denotes the initial mass and $q_c = M_c/M$ is the final fractional core mass, M_c being the mass of the hydrogen-exhausted core and $M = 20 M_{\odot}$, the final mass of the progenitor. The asterisks indicate the parameters of the models just before the supernova event. For comparison, the final parameters of all three cases are shown in each panel.

wards, while the supernova progenitor undergoes advanced evolutionary phases, the neutron star will accrete a portion of any stellar wind emitted by the progenitor. Using the Bondi-Hoyle model²⁰ for wind accretion, we find that the expected X-ray luminosity, $L_X \approx 1 \times 10^{34}$ ergs s⁻¹, is well below the upper limit for the X-ray luminosity in the direction of SN1987A obtained from archival Einstein data²¹ ($L_X \leq 9 \times 10^{34}$ ergs s⁻¹).

More than half the total mass of the binary system will have been ejected in the observed supernova explosion, and the system will now be unbound. The older neutron star (the collapsed remnant of the original primary) may become detectable within the next few years, either as a radio pulsar or as an X-ray source powered by the accretion of material from the supernova remnant.

The principal advantage of either of the above scenarios over single-star models for SN1987A is that a blue progenitor is a natural consequence of conventional binary stellar evolution theory, and does not depend on *ad hoc* assumptions or modifications of the input physics. Moreover, our binary model may account for many of the anomalous features of SN1987A (see also ref. 2). In particular, the natural structure of our pre-supernova model is sufficiently similar to those of single-star models^{15,22,23} that we can expect fits to the supernova light curve entirely as good as those achieved in the single-star models. The asymmetric expansion of the ejecta, as inferred from polarization measurements and speckle interferometry^{24,25}, also has a natural explanation in our model, as we expect the progenitor to be rotationally flattened as a result of the spin-up it experienced during the mass accretion phase (see also R. A. Chevalier, preprint). Moreover, the relativistic-jet models²⁶ that have been invoked to explain the 'mystery spot'²⁷ seem to require a close-binary companion, and the existence of a nearby companion star may account for the observed variability²⁸ of the soft X-ray emission (ref. 29 and R. McCray, personal communication).

Our model also provides a plausible mechanism to explain the anomalously high abundance of barium and other s-process elements in the ejecta³⁰. If the companion passed through a helium-giant phase³¹, s-processed material might have been dredged up³² and subsequently transferred to the progenitor by Roche-lobe overflow. Alternatively, mass transfer might spin up the progenitor nearly to break-up velocity, in which case meridional currents might overcome the stabilizing effect of chemical gradients³³ and produce large-scale mixing of the envelope, including the dredge-up of s-processed material (see also refs 34 and 35). We note that there is no known mechanism to account for this anomaly in single-star scenarios for SN1987A and that all known barium stars are found in binary systems³⁶.

Future observations of the supernova and its environment (such as the nitrogen-rich circumstellar material³⁷) will help to distinguish between the single star and the binary models. However, the ultimate test will be the reappearance of the predicted companion star, be it a relatively normal star, a neutron star, or Sk-69°202 itself. We should know within a few years whether SN1987A was a member of a binary system and, if so, which of the proposed binary models is correct. □

14. Iben Jr, I. *Astrophys. J.* **304**, 201–216 (1986).
15. Nomoto, K. & Shigeyama, T. *Proc. ESO Workshop on Supernova 1987A* (ed. Danziger, I. J.) (in the press).
16. Barkat, Z. & Wheeler, J. C. *Astrophys. J.* **332**, 247–260 (1988).
17. Kippenhahn, R. & Weigert, A. Z. *Astrophys.* **65**, 251–273 (1967).
18. Lauterborn, D. *Astr. Astrophys.* **7**, 150–159 (1970).
19. Tutukov, A. V. & Yungelson, L. R. in *IAU Symp. 88: Close Binary Stars: Observations and Interpretations* (eds Plavec, J. M., Popper, D. M. & Ulrich, R. K.) 15–22 (Reidel, Dordrecht, 1980).
20. Bondi, H. & Hoyle, F. *Mon. Not. R. astr. Soc.* **104**, 273–282 (1944).
21. Harnden Jr, F. R., & Seward, F. D. *Proc. 4th George Mason Workshop on Supernova 1987A* (ed. Kafatos, M.) (Cambridge University Press, Cambridge, in the press).
22. Arnett, W. D. *Astrophys. J.* **319**, 136–142 (1987).
23. Woosley, S. E., Pinto, P. A. & Ensman, L. *Astrophys. J.* **324**, 466–489 (1988).
24. Cropper, M. et al. *Mon. Not. R. astr. Soc.* **231**, 695–722 (1988).
25. Karovska, M. et al. *IAU Circ. No. 4604* (1988).
26. Rees, M. J. *Nature* **328**, 207 (1987).
27. Nisenson, P., Papaliolios, C., Karovska, M. & Noyes, R. *Astrophys. J.* **320**, L15–L18 (1987).
28. Dotani, T. et al. *Nature* **330**, 230–231 (1987).
29. Fabian, A. C. & Rees, M. J. *Nature* **335**, 50–51 (1988).
30. Williams, R. E. *Astrophys. J.* **320**, L117–L120 (1987).
31. Habets, G. M. H. J. *Astr. Astrophys.* **167**, 61–76 (1986).
32. Iben Jr, I. & Tutukov, A. V. *Astrophys. J. Suppl.* **58**, 661–710 (1985).
33. Mestel, L. in *Stellar Structure* (eds Aller, L. H. & McLaughlin, D. B.) Vol. 8 465–497 (University of Chicago Press, Chicago, 1965).
34. Saio, H., Kato, M. & Nomoto, K. *Astrophys. J.* **331**, 388–393 (1988).
35. Weiss, A., Hillebrandt, W. & Truran, J. W. *Astr. Astrophys.* **197**, L11–L14 (1988).
36. McClure, R. D. *Astrophys. J.* **268**, 264–273 (1983).
37. Fransson, C. et al. *Astrophys. J.* **336**, 429–441 (1989).

ACKNOWLEDGEMENTS. This work was supported in part by the National Aeronautics and Space Administration. We are grateful to L. Nelson for providing us with a copy of his stellar evolution code and for many helpful discussions concerning the modification and implementation of the code, and we thank D. VandenBerg for supplying us with his opacity tables.

Stabilization of lamellar oil–water liquid crystals by surfactant/co-surfactant monolayers

L. F. Braganza*, M. Dubois† & J. Tabony†‡

* Institute Laue–Langevin, 156X, 38042 Grenoble Cedex, France

† Département de Physico-Chimie, CEN Saclay, Gif-sur-Yvette, France

LIQUID crystals are divided into two main classes, thermotropic and lyotropic. Thermotropic liquid crystals are formed by melting, whereas lyotropic liquid crystals arise from the association of molecules, such as soap and water, that in general are not in themselves liquid crystalline. Thermotropic liquid crystals are used for liquid-crystal displays; lyotropic liquid crystals occur in living cells. Here we report a novel sequence of lyotropic liquid crystals comprising alternate layers of oil and water whose thickness varies linearly with the relative proportions of oil and water, and we have determined their structure using neutron diffraction methods. The oil and water layers are separated and stabilized by a monolayer film of surfactant and co-surfactant. The individual layers are typically a hundred ångströms or more in thickness, and total lamellar spacings of up to 1,000 Å were observed. This behaviour is difficult to describe in terms of the theories of colloid stability currently used to describe lyotropic liquid crystals. An understanding of the self-organization of such systems over such large distances would elucidate how long-range liquid-crystalline ordering arises in living cells. Moreover, thermotropic liquid crystals are expensive and chemically relatively unstable, and lamellar mesophases of the lyotropic type described here could lead to inexpensive, chemically stable liquid-crystalline materials suitable for industrial application.

Lamellar mesophases of surfactant/water liquid crystals¹ are well characterized structures comprising alternate surfactant bilayer sheets and water layers about 30 Å thick. Diluting these mesophases rarely results in lamellar repeat spacing of more than 100 Å; instead, either a micellar phase is formed or phase separation occurs. Dispersions of oil and water stabilized by surfactant and co-surfactant frequently form fluid isotropic microemulsions². Under some circumstances, however, such

‡ Present address: Département de Recherche Fondamentale, CEN Grenoble, 85X Grenoble Cedex, France.

Received 17 October 1988; accepted 18 January 1989.

1. Fabian, A. C., Rees, M. J., van den Heuvel, E. P. J. & van Paradijs, J. *Nature* **328**, 323–324 (1987).
2. Joss, P. C., Podsiadlowski, Ph., Hsu, J. J. L. & Rappaport, S. *Nature* **331**, 237–240 (1988).
3. Walborn, N. R., Lasker, B. M., Laidler, V. G. & Chu, Y.-H. *Astrophys. J.* **321**, L41–L44 (1987).
4. Falk, S. W. & Arnett, W. D. *Astrophys. J. Suppl.* **33**, 515–562 (1977).
5. Woosley, S. E. & Weaver, T. A. in *Proc. 5th Morand Astrophys. Conf.: Nucleosynthesis and its Implications on Nuclear and Particle Physics* (eds Audouze, J. & Mathieu, N.) 145–166 (Reidel, Dordrecht, 1985).
6. Woosley, S. E. *Astrophys. J.* **330**, 218–253 (1988).
7. Abt, H. A. & Levy, S. G. *Astrophys. J. Suppl.* **36**, 241–258 (1978).
8. Renzini, A. in *IAU Symp. 105: Observational Tests of Stellar Evolution Theory* (eds Maeder, A. & Renzini, A.) 21–40 (Reidel, Dordrecht, 1984).
9. Applegate, J. H. *Astrophys. J.* **329**, 803–807 (1988).
10. Höppner, W. & Weigert, A. *Astr. Astrophys.* **25**, 99–103 (1973).
11. Weiss, A. *Astr. Astrophys.* **127**, 411–412 (1983).
12. Maeder, A. in *IAU Symp. 105: Observational Tests of Stellar Evolution Theory* (eds Maeder, A. & Renzini, A.) 299–319 (Reidel, Dordrecht, 1984).
13. Kippenhahn, R., Weigert, A. & Hofmeister, E. in *Methods in Computational Physics* Vol. 7 (eds Alder, B., Fernbach, S. & Rothenberg, M.) 129–190 (Academic, New York, 1967).

mixtures can give rise to liquid-crystalline phases that have lamellar, hexagonal or cubic structures³⁻⁵. We have studied the lamellar mesophases prepared from tetradecyltrimethylammonium bromide (TTAB, $C_{14}H_{29}N^+(CH_3)_3Br^-$) as surfactant, pentanol as co-surfactant, cyclohexane and water. Three series of compositions were studied: in series A and B, the amounts of surfactant and co-surfactant were constant, as was the total volume of cyclohexane and water, but the relative volume fraction of cyclohexane to water was varied. In series C, the lamellar phase was gradually diluted with an aqueous solution of 1% by weight of sodium bromide pre-saturated with pentanol and cyclohexane. The compositions were as follows: (A) 0.70 g TTAB, 0.38 ml pentanol, 4.0 ml of water and cyclohexane; (B) 0.29 g TTAB, 0.33 ml pentanol, 4.2 ml of water and cyclohexane; (C) TTAB/pentanol/cyclohexane in the weight ratio 1:0.56:0.55 in a 1% by weight solution of NaBr pre-saturated with cyclohexane and pentanol. For composition C the overall TTAB concentration varied from that used in composition B down to one eighth of this value.

In the absence of cyclohexane, composition A formed a birefringent phase which persisted on replacement of water by oil for oil-to-water volume fractions of 90% cyclohexane/10% water. Composition B was not liquid crystalline in the absence of oil but formed a birefringent phase as soon as a small amount of cyclohexane was added. The birefringent phase was stable for samples containing up to 70% cyclohexane/30% water, after which phase separation occurred into optically birefringent and isotropic phases. For surfactant concentrations much less than those used in series B, the oil-in-water lamellar phases became unstable. Stable lamellar phases down to much lower concentrations were obtained by using an aqueous solution of NaBr instead of water.

All the birefringent phases gave neutron diffraction patterns consistent with a lamellar structure, and up to three diffraction orders were observed. For compositions A and B, the overall repeat spacings were 130 Å and 230 Å respectively. The lamellar mesophases orientate in magnetic fields; Fig. 1 shows the neutron intensity distribution over a two-dimensional detector from a typical orientated sample. Neutron-scattering-length densities of protonated and deuterated materials are of very different magnitude. Hence, in multicomponent systems, selective deuteration of different components can be used to highlight

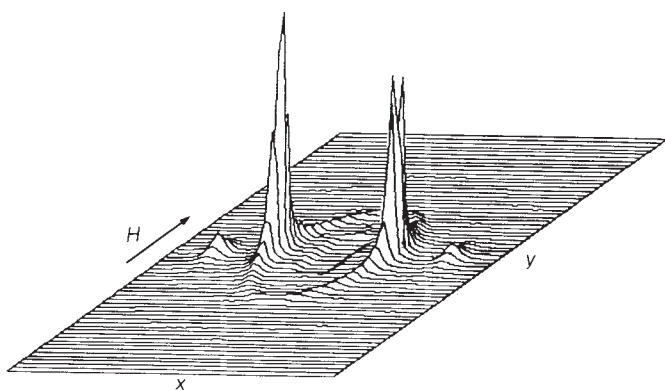


FIG. 1 Neutron scattering pattern on a two-dimensional 64 × 64 cm detector for an orientated mesophase with an oil-to-water volume fraction of 45% cyclohexane/55% D_2O . The sample composition was 0.70 g tetradecyltrimethylammonium bromide, 0.38 ml pentanol, 1.80 ml cyclohexane and 2.20 ml D_2O . All measurements were done on the D17 diffractometer¹⁷ at the Institut Laue-Langevin, Grenoble, France. The pattern was recorded in 20 min using a sample-to-detector distance of 1.40 m and a wavelength of 9.0 Å. The neutron beam, stopped by a rectangular cadmium plate after transmission through the sample, is centred on the detector; two reflections on either side correspond to the first and second lamellar diffraction orders. Orientation was achieved by cooling the sample from 60 °C to 20 °C in a magnetic field of 6.5 tesla over 4 h. Shown in the figure is the direction of the magnetic field, H , with respect to the detector.

specific features of the structure. Figure 2 shows a Fourier profile, derived from the phased structure factors using standard procedures⁶, for a mesophase of composition A. The profile corresponds to a low-resolution image of the variation in neutron-scattering-length density normal to the lamellae. Both the water and cyclohexane were deuterated; the maxima correspond to the layers of deuterated oil and water, and minima correspond to the protonated surfactant/co-surfactant layers. Fourier profiles obtained for samples containing either surfactant or pentanol as the only protonated components were nearly identical to those in Fig. 2, indicating that these two components share a common location within the structure. Most of the pentanol molecules are therefore intercalated between the surfactant chains. Deuteration of oil or water reduces alternate maxima, corresponding to either the protonated oil or water layers, to the level of the protonated surfactant/co-surfactant minima. For samples consisting of D_2O , deuterated cyclohexane, deuterated pentanol and $C_{14}D_{29}N(CH_3)_3Br^-$, the scattering arises solely from the surfactant polar heads. The thickness of the surfactant/co-surfactant layer was determined as ~ 17 Å, equal to half of the bilayer thickness derived from samples containing no cyclohexane. These results show that the lamellar structure comprises alternating layers of cyclohexane and water separated and stabilized by a monolayer film of surfactant and co-surfactant. As the cyclohexane-to-water volume fraction increases, there is no change in the total repeat spacing but the cyclohexane layer thickens at the expense of the water region. Shown in Fig. 3 are plots of the overall spacing and the thicknesses of cyclohexane and water layers versus the cyclohexane-to-water volume fraction. The cyclohexane layer thickness increases almost linearly with oil content, and a corresponding decrease occurs in the D_2O layers, until for composition A at 90% cyclohexane/10% D_2O the D_2O layer thickness is ~ 10 Å. Similar results were obtained for composition B, for which the overall repeat-unit thickness was 230 Å, and the cyclohexane and D_2O layer thicknesses, just before phase separation (at an oil-to-water volume fraction of greater than 70% cyclohexane/30% D_2O), were ~ 110 Å and ~ 80 Å, respectively. For

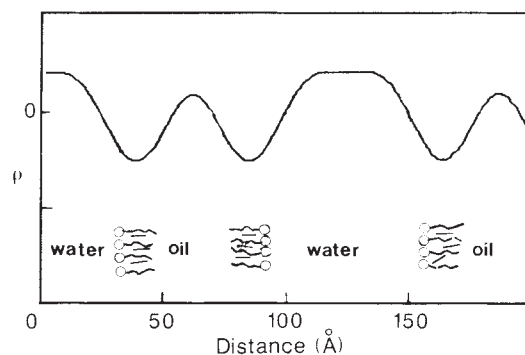


FIG. 2 Scattering-length density profile at 20.0 ± 0.2 °C for a mesophase composed of 0.70 g tetradecyltrimethylammonium bromide, 0.38 ml pentanol, 1.4 ml deuterated cyclohexane and 2.6 ml D_2O . The repeat spacing is 125 ± 2 Å. Also shown is a schematic drawing of the structure, with pentanol molecules represented as dark rods intercalated between surfactant molecules. The Fourier profiles were calculated from $\rho(x) = \pm |F_n| \cos(2\pi nx/d)$, where $\rho(x)$ is the coherent scattering amplitude density, $|F_n|$ is the structure-factor modulus of the n th order diffraction peak and d is the repeat spacing. Because preparations with different oil-to-water volume ratios gave different degrees of magnetic orientation, non-orientated samples were used, and structure factors were calculated from the diffraction intensities I_n according to $|F_n| = (\sin 2\theta_n \sin \theta_n)^{1/2}$, where $(\sin 2\theta_n \sin \theta_n)$ is the Lorentz factor as applied to randomly orientated or powder samples. Phases—equal to either zero or π for lamellar structures—were determined by substituting deuterated for protonated molecular components and by swelling the structure (by changing the temperature). Diffraction patterns were recorded typically in 60 min on the diffractometer D16 at the Institut Laue-Langevin¹¹. Statistical errors in the structure factors were better than $\pm 1\%$ for the strong reflections and a few per cent for the weak peaks.

composition C (1% NaBr solution), samples containing much less surfactant and co-surfactant were strongly birefringent, and the observed lamellar spacings ranged up to 1,000 Å.

In most surfactant-water mixtures, lamellar phases do not arise until concentrations of ~50% by weight are attained, such that the thicknesses of the surfactant bilayer and the water layer are comparable (~30 Å). The addition of a co-surfactant such as pentanol promotes the formation of lamellar phases at lower surfactant concentrations, giving higher spacings. For example, in the absence of cyclohexane, composition A forms a lamellar phase which persists down to surfactant concentrations of ~10% by weight. At lower concentrations, micelles are formed. This behaviour can be explained in the following manner. The pentanol molecules are intercalated between the surfactant hydrocarbon chains, reducing the curvature of the micelle-water interface until flat surfaces are formed and a lamellar packing results. The addition of both co-surfactant and oil favours the formation of a lamellar phase for surfactant concentrations even lower than 10% by weight; by adding KBr, lamellar phases formed at less than 1% of surfactant. Once a flat monolayer has been formed, the lamellar structure may be retained whilst significantly modifying the relative proportions of oil and water. There has been considerable debate concerning the structure of microemulsions containing equal volumes of oil and water⁷⁻⁹. One possibility comprises bicontinuous structures of cubic symmetry in which the local surface curvature is non-zero but the total curvature is very low^{3,7-9}. Another possibility is the lamellar structures such as those reported here. It should be possible to effect a transition between the two types of structure by varying the co-surfactant content and thus modifying the local interfacial curvature.

These arguments based on interfacial curvature may help to explain why these lamellar structures arise. However, the nature of the forces operative in this system, for which aqueous-based or oil-rich mesophases can be prepared with neither loss of the lamellar structure nor a change in the repeat structure, remains to be determined. The stability of lamellar mesophases relative to other structures may be understood classically in terms of repulsive electrostatic interactions and attractive van der Waals forces^{10,11}. Both of these contributions are strongly modified by replacing water by oil and both decay rapidly with increasing distance. Electrostatic interactions decay exponentially with the inverse Debye length. The Debye lengths calculated for the systems with no added NaBr varied from ~5 to 10 Å, and were less than 10 Å for the systems with added NaBr. Classical theories of electrostatics do not seem to account for the observed

behaviour; in particular, decreasing the Debye length by adding salt results in a stabilization of the lamellar phase to longer layer spacings. This, and the existence of a lamellar structure over a wide range of oil-to-water volume ratios, suggests that additional long-range forces might be operating. An interfacial-fluctuation-driven mechanism¹² has been suggested in this respect, but the applicability of this mechanism has not been established, and it is difficult to see how it can explain recent results obtained on clay/water systems¹³. For aqueous layers between hydrophobic surfaces, Israelachvili and Pashley *et al.*^{14,15} deduced the presence of strong long-range attractive hydrophobic forces, which are difficult to explain in classical terms. If such long-range forces exist, then they would be of considerable importance in many areas of molecular biology and colloid science. The lamellar mesophases presented here may be suitable systems in which to probe their nature, by, for example, force-distance measurements. These lamellar oil-water mesophases may find several applications¹⁶. In particular, lyotropic liquid crystals with a range of tailored properties may be prepared by the inclusion of appropriate compounds or colloidal materials into the aqueous, hydrocarbon or interfacial regions. Structural studies of the type reported here are an essential step towards the development and exploitation of this novel technology. □

Received 17 October 1988; accepted 6 February 1989.

1. Ekwall, P. in *Advances in Liquid Crystals* (ed. Brown, G. H.) Vol. 1, Ch. 1 (Academic, New York, 1971).
2. Prince, L. M. *Microemulsions: Theory and Practice* (Academic, New York, 1977).
3. Tabony, J. *Nature* **319**, 400 (1986).
4. Tabony, J. *Nature* **320**, 338-339 (1986).
5. Roux, D. & Safinya, C. R. in *Physics of Amphiphile Layers* (eds Meunier, J., Boccaro, N. & Langevin, D.) Springer Proc. in Physics **21**, 138 (1987).
6. Franks, N. J. *molec. Biol.* **100**, 345-358 (1976).
7. Scriven, L. E. *Nature* **263**, 123-125 (1976).
8. Meunier, J., Boccaro, N. & Langevin, D. (eds) *Physics of Amphiphile Layers* Springer Proc. in Physics **21** (1987).
9. De Geyer, A. & Tabony, J. in *Physics of Amphiphile Layers* (eds Meunier, J., Boccaro, N. & Langevin, D.) Springer Proc. in Physics **21**, 372 (1987).
10. Derjaguin, B. & Landau, L. D. *Acta phys. Chim. U.S.S.R.* **14**, 633 (1941).
11. Verwey, E. J. W. & Overbeek, J. Th. G. *Theory of Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
12. Helfrich, W. Z. *Naturf.* **33a**, 305-315 (1978).
13. Braganza, L. F., Crawford, R. J., Smalley, M. V. & Thomas, R. K. *Clays Clay Miner.* (in the press).
14. Pashley, R. M., McGuigan, P. M., Ninham, B. W. & Fennell Evans, D. *Science* **229**, 1088-1089 (1985).
15. Israelachvili, J. N. & Pashley, R. M. *Nature* **300**, 341-342 (1982).
16. Tabony, J. *French Patents* 87 40 25 017, 87 40 25 018 (1987).
17. *Neutron Beam Facilities Available for Users* (Institut Laue-Langevin, Grenoble, 1981).

Atmospheric ¹⁴C and century-scale solar oscillations

Minze Stuiver & Thomas F. Braziunas

Department of Geological Sciences and Quaternary Research Center, University of Washington, Seattle, Washington 98195, USA

THE solar-wind plasma in our interplanetary space deflects part of the Earth-bound cosmic-ray flux through its magnetic interaction with the electrically charged incoming particles. Because changes in the magnetic properties of the plasma originate at the Sun's surface, the cosmic-ray flux arriving at Earth therefore depends on the changing surface conditions of the Sun. Consequently, by monitoring the variable production rate of cosmogenic isotopes (such as ¹⁴C) in our atmosphere, a time history of changing conditions of the Sun's surface can be obtained. Trees, through carbon dioxide assimilation, lay down a ¹⁴C record which provides clues towards the causes underlying ¹⁴C production-rate changes. Here we show from maximum-entropy spectral analysis of a 9,600-yr high-precision ¹⁴C chronology that changes occur in the Sun's convective zone with a fundamental oscillatory mode of about $2.4 \times 10^{-3} \text{ yr}^{-1}$ (420-yr period), and we also identify several harmonics. Previous searches¹⁻³ for cyclicity in the atmospheric ¹⁴C record have yielded periods near 140 and 200 yr. We discuss the implications of a longer and more precise ¹⁴C record.

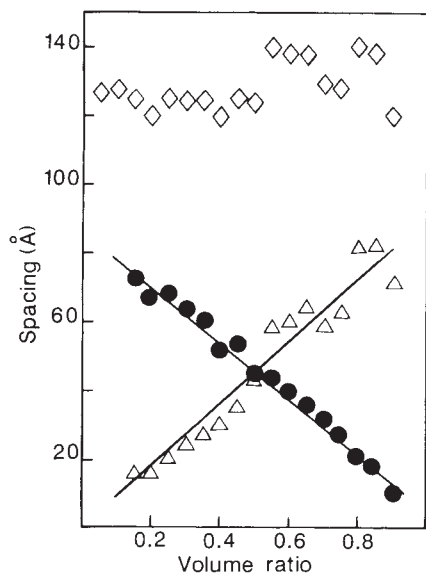


FIG. 3 Plots of repeat spacing (◇), and water (●) and cyclohexane (△) layer thicknesses, as a function of oil-to-water volume ratio, for composition A.

Determinations for this study were made by the precise counting of the ^{14}C radioactivity of large samples, the sample activity being compared with the absolute National Bureau of Standards (NBS) oxalic acid standard activity. For tree-ring materials of known age a correction for ^{14}C decay is applied. The ^{14}C results are also normalized on a fixed $\delta^{13}\text{C}$ value of -25% relative to the PDB standard

$$\left(\delta^{13}\text{C} = \left[\frac{(^{13}\text{C}/^{12}\text{C})_{\text{sample}}}{(^{13}\text{C}/^{12}\text{C})_{\text{PDB}}} - 1 \right] 1,000\% \right)$$

The results are expressed as $\Delta^{14}\text{C}$ (ref. 4), which represents the ^{13}C -normalized ^{14}C activity at the time of formation, expressed as the deviation in parts per thousand from the oxalic acid standard activity.

A compilation of the tree-ring ^{14}C data reported in ref. 5 yielded the basic 9,600-yr bi-decadal $\Delta^{14}\text{C}$ record used here. The data of the Seattle^{6,7} and Belfast^{7,8} laboratories extend back to 5,200 yr BC, whereas a mixture of results from Seattle⁹, La Jolla¹⁰, Tucson¹¹ and Heidelberg¹² was used for the interval 5200–7750 BC. The Seattle–Belfast portions of the data set have a typical standard deviation of 1.5–2‰ for bi-decadal wood. Such precision includes overall reproducibility as well as the Poisson error in the number of counts. For pre-5200 BC data, the average standard deviation derived from the scatter of the results reported within 20-yr blocks is 5‰.

The long-term $\Delta^{14}\text{C}$ record shows a broad 10% decline towards the present, which is usually attributed to the influence either of an increasing geomagnetic dipole moment on ^{14}C production rate or of changing carbon-reservoir parameters (that is, climatic conditions) on atmospheric $\Delta^{14}\text{C}$. Because we focus attention here on solar modulation of the cosmic-ray flux, the long-term trend was removed by subtracting a spline approximating a 400-yr moving average. Our contention that most, if not all, of the century-scale oscillations in the residual $\Delta^{14}\text{C}$ record (Fig. 1) are heliomagnetic in origin is supported by calculations of the magnitude of the expected solar-induced $\Delta^{14}\text{C}$ variations¹³, by the coincidence of the $\Delta^{14}\text{C}$ maximum near AD 1700 with the lack of sunspots experienced during the AD 1654–1714 Maunder minimum^{13,14}, by the compatibility of the ^{14}C history of the last millennium with auroral evidence and naked-eye sunspot observations^{13,15} and by the similarity (in timing as well as magnitude) of the ^{10}Be cosmogenic record in ice cores and the atmospheric $\Delta^{14}\text{C}$ record¹⁶.

The $\Delta^{14}\text{C}$ oscillations of a few per cent in Fig. 1 have approximately equal rates of $\Delta^{14}\text{C}$ increase and decrease. They differ slightly in length: the Maunder oscillations have a period of ~ 180 yr and the Spörer oscillations last ~ 40 yr longer¹⁷. These oscillations are denoted M and S in Fig. 1. Nine of the Maunder-type oscillations and eight of the Spörer variety are found in the 9,600-yr record. Episodes of triple oscillations, discussed later, are labelled T_1 , T_2 , T_3 and T_4 .

The atmospheric ^{14}C signal results from frequency-dependent attenuation of variations in ^{14}C production rate, Q , and so is a less direct record of heliomagnetic change than is Q itself. We calculated Q from the observed $\Delta^{14}\text{C}$ record by carbon-reservoir modelling^{13,17}, and determined the spectral distribution of the entire 9,600-yr Q record, without removing any trends. From a Fourier analysis, spectral power exceeding the 2σ significance level¹⁸ was found at periods of 230–200, 150, 88 and 57 yr. The maximum entropy method (MEM) or autoregressive (AR) model, using the Burg¹⁹ or FABNE²⁰ algorithms and AR order 20, yields spectral power for similar periods (Fig. 2). The minimum final-prediction error (FPE) was the selection criterion for AR order 20 (ref. 21). Virtually the same set of cycles (425, 227, 148, 85, 67, 58 and 45 yr) is found through the MEM approach (AR order 20) after de-trending Q in the same manner as shown in Fig. 1 for the atmospheric $\Delta^{14}\text{C}$ record.

A set of harmonics underlies the periods in Fig. 2, as demonstrated by the inset, in which the hyperbola represents the

harmonics of the 420-yr period. The fourth harmonic, with a calculated period of 105 yr, is not present in the entire spectrum of Fig. 2, but on subdividing the record into four equal parts we find 110- and 107-yr cycles in, respectively, the 5350 BC–2970 BC and 570 BC–AD 1830 portions.

The cycles with 67-, 52- and 45-yr periods contain about one twentieth of the spectral power of the longer cycles in the Fourier spectrum. They may be related to the non-sinusoidal shape of the slower cycles. For instance, a square wave with a 420-yr period generates harmonics in our fast-Fourier-transform spectrum at about one twentieth of the strength of the 420-yr peak at periods of 139, 84, 60 and 47 yr.

The harmonics shown in Fig. 2 express the average oscillatory properties of the 9,600-yr record. A least-squares procedure was used to determine phase relationships and relative amplitudes for each of the principal oscillations (420, 218, 143 yr) generated by the MEM method. The iterative procedure involved testing combinations of phases and amplitudes for periods near each of the principal oscillation periods until a best fit with the observed record was achieved. The relative phases of the solar

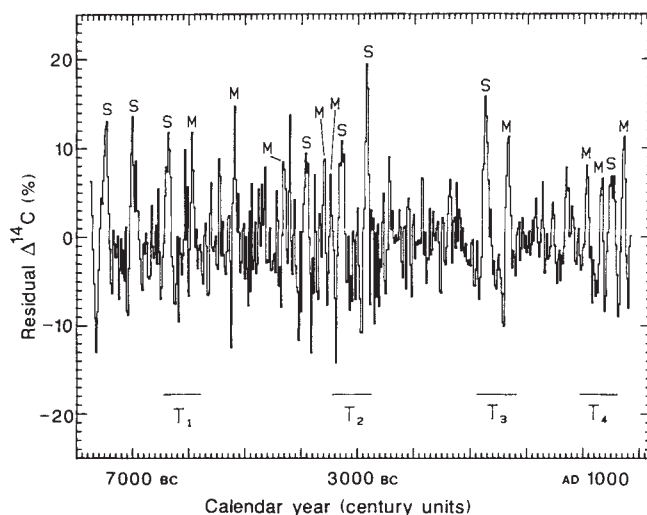


FIG. 1 Residual atmospheric $\Delta^{14}\text{C}$, obtained by deducting the long-term $\Delta^{14}\text{C}$ trend. M and S denote Maunder- and Spörer-type events. Episodes of triple oscillations (T_1 – T_4) are discussed in the text.

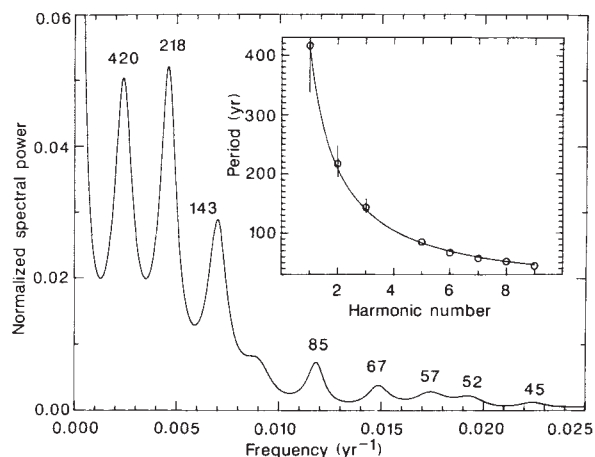


FIG. 2 MEM power spectrum (AR=20) of ^{14}C production rate Q . The periods are listed with the individual peaks. The hyperbola in the inset shows the harmonics expected for a fundamental frequency of $1/420 \text{ yr}^{-1}$. The observed periods are plotted in the inset with vertical bars representing one standard deviation (error bars smaller than the symbol size are not shown).

periodicities were assumed to remain unchanged for the entire 9,600-yr record. The periods obtained in this manner differ slightly from those in the raw MEM analysis.

Improved coherence is obtained for different sections of the record when two variants of the average (420, 218, 143 yr) mode are considered. Sections of the record (T_1 , T_2 , T_3 and T_4 in Fig. 1), each of which contains at least two of the Maunder- and Spörer-type digressions, show clear similarities when viewed on an amplified timescale¹⁷. A best fit to these segments, retaining their relative timing, yields equation (1), whereas the remaining sections are best represented by equation (2):

$$\frac{\Delta Q}{Q_t}(\%) = 8.8 \cos\left(\frac{2\pi}{435}t + 1.5\pi\right) + 11.2 \cos\left(\frac{2\pi}{207}t + 1.6\pi\right) + 5.6 \cos\left(\frac{2\pi}{148}t + 0.5\pi\right) \quad (1)$$

$$\frac{\Delta Q}{Q_t}(\%) = 2.8 \cos\left(\frac{2\pi}{456}t + \pi\right) + 4.3 \cos\left(\frac{2\pi}{256}t + 0.8\pi\right) + 3.4 \cos\left(\frac{2\pi}{143}t + 0.3\pi\right) \quad (2)$$

with $\Delta Q/Q_t$ being the percentage deviation of Q from the splined long-term trend Q_t , and t being time in calendar years (negative for dates BC).

In Fig. 3b the solid curve shows the correlation coefficients between sliding 700-yr portions of equation (1) (435, 207, 148 variant) and the ^{14}C production record, and the dotted curve gives the corresponding correlation for equation (2) (456, 256, 143 variant). The first variant generates the basic pattern of the four triplets T_1 – T_4 without changing phase or amplitude of the sine waves (Fig. 3a).

The discussion so far has centred on the periodicities generated by the MEM method using moderate AR numbers. By increasing the AR order or using Fourier power spectra, additional periods become important. For instance, the single 420-yr peak (Fig. 2) splits into 504-, 355- and 299-yr peaks in the Fourier analysis of the complete de-trended Q record. Spectral properties also vary to some extent for different parts of the record (for example, in the 570 BC–AD 1830 quarter section a pronounced 127-yr peak, absent in the other sections, is observed).

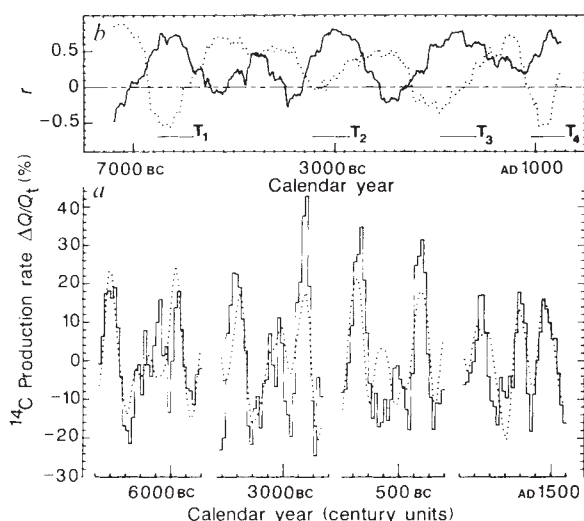


FIG. 3 a, Tree-ring-derived changes in ^{14}C production rate Q of the triplet episodes T_1 – T_4 (solid line), compared with the curve generated from equation (1) (dotted line). Triplet intervals are 6480–5800 BC, 3420–2740 BC, 880–200 BC and AD 920–1600. b, Correlation coefficient r of a 700-yr sliding portion of the ^{14}C production rate with curves generated from either equation (1) (solid line) or equation (2) (dotted line). The 700-yr sliding window results in a loss of 350 yr at either end of the curves.

Adding frequencies to equation (1) and using variable amplitude or phase shift would improve the correlation, but our goal here was to explain the major features of the ^{14}C record in the least complicated manner.

Although we have argued that the main century-scale atmospheric ^{14}C oscillations are best explained in terms of variations in ^{14}C production rate induced by solar change, we have also considered the oceanic changes that would be required to produce these atmospheric patterns. Atmospheric $\Delta^{14}\text{C}$ can increase either by a reduction in the CO_2 gas exchange rate at the air–sea interface or by a reduction in the rate of upwelling of ^{14}C -deficient deep ocean water. Carbon-reservoir modelling indicates that global wind speeds would have to be reduced to two-thirds of the current average to create a Maunder $\Delta^{14}\text{C}$ event; alternatively, mixing rates in the global ocean would have to fall back rapidly to half the long-term average. Such changes are large, and there are no proxy data to support such extreme behaviour. Thus we consider that the $\Delta^{14}\text{C}$ oscillations are best explained by solar-wind forcing.

Evidence of global solar change is provided by the 22-yr Hale (sunspot) cycle, from which a 90-yr (Gleissberg) periodicity has been deduced. Analysis of the number of aurorae reported per decade gives periodicities of 88 and 130 yr (refs 23, 24). The 88-yr periodicity is near the 85–87-yr period derived from the ^{14}C data (Fig. 2), whereas the 130-yr period is close to the 127-yr period found for the 570 BC–AD 1830 interval. Our present analysis adds periodicities of about 140, 220 and 420 yr to these observations of solar patterns.

Solar-constant changes of up to 0.1% have been measured for the recent part of the Hale cycle^{25,26}. Evidently the changes that result in sunspot and cosmic-ray modulation also induce changes in the solar constant. The longer periodicities of the ^{14}C production record may also relate to such solar-constant change, and these changes may be larger in magnitude. If so, a relationship between climate and solar variability, as evidenced in the ^{14}C record, is to be expected. Conflicting evidence exists for such a relationship²⁷. Strong evidence for a correlation was given by Sonett and Suess²² for a tree (bristlecone pine) ring-width record, for which periodicities of, for example, 114 and 208 yr were found.

Applying AR techniques to ten climate records (tree-rings and glacier fluctuations) reported by Röthlisberger²⁸, we find that none of the spectra contain significant power in the range 200–250 yr. However, there is power in the 123–143-yr interval for six of the records, in the 102–104-yr range for three, and near 88 yr for two. In our previous study¹ of climate records covering only one millennium we found no convincing evidence for a Sun–climate relationship. The use of the much longer Röthlisberger records together with the ^{14}C analysis reopens this question, and in particular there may be a Sun–climate relationship for the third harmonic (140 yr). □

Received 21 November 1988; accepted 9 February 1989.

1. Stuiver, M. *Nature* **286**, 868–871 (1980).
2. Neftel, A., Oeschger, H. & Suess, H. E. *Earth planet. Sci. Lett.* **56**, 127–147 (1981).
3. Sonett, C. P. *Rev. Geophys. Space Phys.* **22**, 239–254 (1984).
4. Stuiver, M. & Polach, H. S. *Radiocarbon* **19**, 355–363 (1977).
5. Stuiver, M. & Kra, R. (eds) *Radiocarbon* **28**, 805–1030 (1986).
6. Stuiver, M. & Pearson, G. W. *Radiocarbon* **28**, 805–838 (1986).
7. Pearson, G. W. & Stuiver, M. *Radiocarbon* **28**, 839–862 (1986).
8. Pearson, G. W., Pilcher, J. R., Baillie, M. G. L., Corbett, D. M. & Qua, F. *Radiocarbon* **28**, 911–934 (1986).
9. Stuiver, M., Kromer, B., Becker, B. & Ferguson, C. W. *Radiocarbon* **28**, 969–979 (1986).
10. Linick, T. W., Suess, H. E. & Becker, B. *Radiocarbon* **27**, 20–32 (1985).
11. Linick, T. W., Long, A., Damon, P. E. & Ferguson, C. W. *Radiocarbon* **28**, 943–953 (1986).
12. Kromer, B. *et al. Radiocarbon* **28**, 954–960 (1986).
13. Stuiver, M. & Quay, P. D. *Science* **207**, 11–19 (1980).
14. Eddy, J. A. *Science* **192**, 1189–1202 (1976).
15. Stuiver, M. & Grootes, P. M. in *The Ancient Sun* (eds Pepin, R. O., Eddy, J. A. & Merrill, R. B.) 165–173 (Pergamon, New York, 1980).
16. Beer, J. *et al. Nature* **331**, 675–679 (1988).
17. Stuiver, M. & Brazunas, T. F. in *Secular Solar and Geomagnetic Variations in the Last 10,000 Years* (eds Stephenson, F. R. & Wolfendale, A. W.) 245–266 (Kluwer, Dordrecht, 1988).
18. Mitchell, J. M. *et al. Tech. Note No. 79*, 33–79 (World Meteorological Organization, No. 195, TP100, Geneva, 1966).
19. Burg, J. P. thesis, Stanford Univ. (1975).
20. Barrodale, I. & Erickson, R. E. *Geophysics* **45**, 433–446 (1980).

21. Akaike, H. *Ann. Inst. Statist. Math.* **21**, 243–247 (1969).
22. Sonett, C. P. & Suess, H. E. *Nature* **307**, 141–143 (1984).
23. Feynman, J. & Fougere, P. F. *J. geophys. Res.* **89**, 3023–3027 (1984).
24. Attolini, M. R., Galli, M. & Nanni, T. in *Secular Solar and Geomagnetic Variations in the Last 10,000 Years* (eds Stephenson, F. R. & Wolfendale, A. W.) 49–68 (Kluwer, Dordrecht, 1988).
25. Schatten, K. H. *Geophys. Res. Lett.* **15**, 121–124 (1988).
26. Wilson, R. C. & Hudson, H. S. *Nature* **332**, 810–812 (1988).
27. Wigley, T. M. L. in *Secular Solar and Geomagnetic Variations in the Last 10,000 Years* (eds Stephenson, F. R. & Wolfendale, A. W.) 209–224 (Kluwer, Dordrecht, 1988).
28. Röhlsberger, F. in *10000 Jahre Gletschergeschichte der Erde* (Sauerländer, Aarau, 1986).

ACKNOWLEDGEMENTS. This work was supported by the NSF. T. L. Saling contributed to the MEM analysis.

Influence of aqueous aluminium and organic acids on measurement of acid neutralizing capacity in surface waters

T. J. Sullivan^{*†}, C. T. Driscoll[†], S. A. Gherini[‡],
R. K. Munson[‡], R. B. Cook[§], D. F. Charles^{||}
& C. P. Yatsko[†]

^{*} Environmental Research Laboratory—Corvallis, 200 SW 35th Street, Corvallis, Oregon 97333, USA

[†] Department of Civil Engineering, Syracuse University, Syracuse, New York 13244, USA

[‡] Tetra Tech, Inc., 3746 Mt Diablo Boulevard, Lafayette, California 94549, USA

[§] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37381, USA

^{||} Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

ACID neutralizing capacity (ANC) is used to quantify the acid-base status of surface waters. Acidic waters have been defined as having ANC values less than zero, and acidification is often quantified by decreases in ANC. Measured and calculated values of ANC generally agree, except for low-ANC waters. These waters, however, are of primary interest in lake-acidification studies. The discrepancy in ANC values is greatest for waters with high concentrations of aluminium and/or dissolved organic carbon (DOC). The discrepancy due to aluminium increases with increasing concentration of dissolved monomeric aluminium (Al_m) and can exceed $50 \mu\text{eq l}^{-1}$ at low pH and high Al_m values. Here we show that this error can be minimized by using a proton reference level for aqueous aluminium of 2+, rather than 0 or 3+, as is done usually. The value of 2+ is consistent with the mean charge on aluminium at the equivalence point of strong-acid titrations. The discrepancy between calculated and measured ANC also increases with increasing DOC concentrations, exceeding $50 \mu\text{eq l}^{-1}$ for waters with DOC concentrations greater than $800 \mu\text{mol C l}^{-1}$ ($1000 \mu\text{mol C l}^{-1} = 12 \text{ mg l}^{-1} \text{ DOC}$). This error can be decreased by titrating to a low consistent pH value (for example, 3.0). This reduces the systematic underestimation of ANC due to curvature in Gran¹ plot analysis, while still allowing accurate measurement of increments in hydrogen-ion concentration. ANC should not be used as a single parameter for characterizing the chemical suitability of surface waters for biota or for assessing the susceptibility of low-ANC waters to acidification by acid deposition.

In regional investigations of acid-base status, ANC has been the principal classification variable². ANC is also widely used by simulation models that predict the response of ecosystems to changes in atmospheric acid deposition^{3–7}. Historical changes in surface-water quality have been evaluated using measured changes in ANC⁸ or estimated by inferring past and present pH and ANC from lake sediment diatom assemblages^{9,10}.

ANC has also been used extensively to assess the susceptibility of surface waters to atmospherically derived strong acids^{11,12}. Waters with high ANC values would generally be considered resistant to changes in pH resulting from changes in deposition acidity. Low-ANC waters may or may not resist changes in pH, depending on the processes supplying ANC in the terrestrial or aquatic environment¹³.

ANC has been used as a primary indicator of a water's suitability for aquatic biota. Other water-quality parameters, however, also need to be considered, including pH (ref. 14), inorganic monomeric aluminium¹⁴, and calcium¹⁵, as well as habitat characteristics (such as the availability of spawning substrate). In low-ANC waters, inorganic monomeric aluminium may constitute a large fraction of the non-hydrogen cations (see below) and thus be of biological significance.

ANC is a measure of the net strong base in solution. It is measured by quantifying the amount of strong acid that must be added to a solution to neutralize this base. The end point of this strong-acid titration would be easily identified except for the presence of weak acids and the relatively small amounts of strong base present in low-ANC waters. Together, these factors obscure the end point. For such systems, the Gran¹ procedure is most commonly used to determine the end point and therefore ANC.

ANC can be calculated by two distinct methods, which have been shown to be mathematically equivalent using the principles of conservation of charge and conservation of mass¹⁶. In one method¹⁷, ANC is calculated as the difference between the sum of the proton (H^+ -ion) acceptors and the sum of the proton donors, relative to a selected proton reference level (discussed below):

$$\text{ANC} = [\text{HCO}_3^-] + 2[\text{CO}_3^{2-}] + [\text{OH}^-] + [\text{other proton acceptors}] - [\text{H}^+] \quad (1)$$

Here, square brackets denote molar concentrations. The other method relates ANC to the total non-hydrogen cation concentrations (M_T ; free and complexed forms); the individual uncomplexed cation charges (z_i) at the equivalence point (the point at which, during titration, the concentration of proton donors equals the concentration of proton acceptors); the total strong-acid anion concentrations (A_T ; free and complexed); and the individual uncomplexed anion charges (z_j), at the equivalence point (general theory¹⁶; practice^{18,19}) using the following relation:

$$\text{ANC} = \sum_i^n |z_i| M_{T_i} - \sum_j^m |z_j| A_{T_j} \quad (2)$$

For aerobic surface waters the first term on the right hand side of equation (2) is typically approximated as

$$\sum_i^n |z_i| M_{T_i} \approx 2[\text{Ca}_T] + 2[\text{Mg}_T] + 2[\text{Mn}_T] + [\text{K}_T] + [\text{Na}_T] + [\text{NH}_4] + x[\text{Al}_T] \quad (3)$$

and the second term is approximated as

$$\sum_j^m |z_j| A_{T_j} \approx 2[\text{SO}_{4T}] + [\text{NO}_{3T}] + [\text{Cl}_T] + [\text{F}_T] \quad (4)$$

The charges z_i and z_j (and thus the concentration multipliers in equations (3) and (4)) are determined by the predominant charges of the uncomplexed constituents at the equivalence point.

For most of the species, there is little uncertainty as to the predominant uncomplexed charge at the equivalence point. For example, the charge of calcium is 2+, and thus the multiplier is 2 in equation (3). However, because of complexation with OH, F and organic ligands, the charge of aluminium, shown as x in equation (3), is not always obvious. Designation of this charge, however, establishes the proton reference level (PRL).

^{*} Present addresses: E&S Environmental Chemistry, Inc., PO Box 609, Corvallis, Oregon 97339, USA and Environmental Research Laboratory—Corvallis, 200 SW 35th Street, Corvallis, Oregon 97333, USA.

PRL constituents do not change ANC but may change pH when added to waters isolated from soils and sediments²⁰. In the literature^{5,18,19}, two PRLs have been used for aluminium, 3+ and 0. These levels have different advantages; the former yields results that are closer to Gran ANC values; the latter eliminates the need to include aluminium in ANC calculations.

Data collected during the regionalized integrated lake-watershed acidification study (RILWAS)^{21,22} from 25 lake-watershed systems in the Adirondack mountains of New York were used

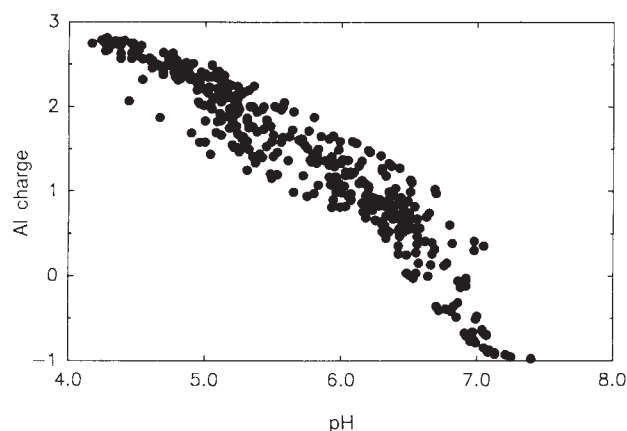


FIG. 1 Mean charge of monomeric inorganic Al versus pH for RILWAS waters. Equivalence point of Gran titrations for dilute waters generally occurs in pH range 4.8 to 5.2. Here the equivalent charge of the aluminium is near 2+.

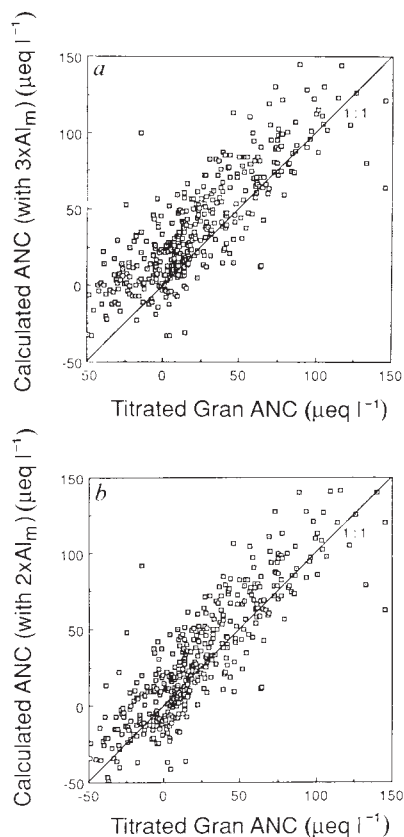


FIG. 2 Calculated ANC versus titrated Gran ANC for RILWAS lakes: Calculated ANC using a, proton reference level for Al_m of 3+ and b, proton reference level of 2+.

to estimate the aluminium PRL. The speciation of aluminium was calculated using the chemical equilibrium model ALCHEMI²³, and the equivalent charge on the aluminium species was determined (Fig. 1). The mean charge on the aluminium increases with decreasing pH. However, over the pH range from 4.8 to 5.2, which corresponds to the equivalence point of dilute waters²⁴, an aluminium charge of 2+ appears more representative than 3+ or zero. Moreover, Gran ANC more closely follows calculated ANC where two times Al_m is used in the calculation for RILWAS lakes (Fig. 2). This is equivalent to the PRL species for Al_m being $Al(OH)^{2+}$ instead of Al^{3+} or $Al(OH)_3^0$.

The difference between calculated and measured Gran ANC values increases as organic-acid concentration, reflected by DOC, increases. This is evident in surface-water data for Meander Lake basin in northern Minnesota (Fig. 3). Meander Lake DOC values average only about $400 \mu\text{mol C l}^{-1}$, but one of the lake's inlets drains a spruce bog and contains DOC values as high as $2,400 \mu\text{mol C l}^{-1}$ (ref. 13). Concentrations of aluminium in Meander waters are very low (mean $Al_m = 1.3 \mu\text{mol l}^{-1}$). The discrepancy between calculated and Gran ANC can be as large as $50 \mu\text{eq l}^{-1}$ at DOC concentrations of $800 \mu\text{mol C l}^{-1}$. A single maximum discrepancy of $180 \mu\text{eq l}^{-1}$ was observed for a sample with $2,400 \mu\text{mol C l}^{-1}$.

To explore the DOC-related discrepancy, Gran analyses were performed on theoretical titration data. The Gran analysis of ANC titration data requires plotting the Gran function F_1 versus titrant volume. F_1 is calculated as follows:

$$F_1 = (V + V_0)10^{-pH} \quad (5)$$

where V is the added titrant volume and V_0 is the original sample volume. This produces a curve like that shown in Fig. 4a. The right-most 'linear' portion of the curve is extrapolated to determine the equivalent volume of titrant, and thus the ANC of the sample.

The Gran analysis procedure was applied to data generated by performing computer-simulated titrations using equilibrium constants associated with the carbonate system¹⁷. In addition, a variety of monoprotic weak acids with varying concentrations and fixed strengths (pK_A values) were added to the solutions, one at a time, to simulate the effects of organic acids ($6.5 \mu\text{eq mg}^{-1}$ DOC, which is close to that observed by others²⁵). The initial ANC of each solution was set to $50 \mu\text{eq l}^{-1}$, and 100 ml aliquots of the solution were titrated with 0.01 M strong acid, in 20 μl increments to pH 3. Table 1 shows the results of the Gran plot analysis using points from pH 3.0 to 3.3 for the extrapolation.

These values suggest that the difference between theoretical and Gran ANC due to the organic acid is <5%, at organic-acid

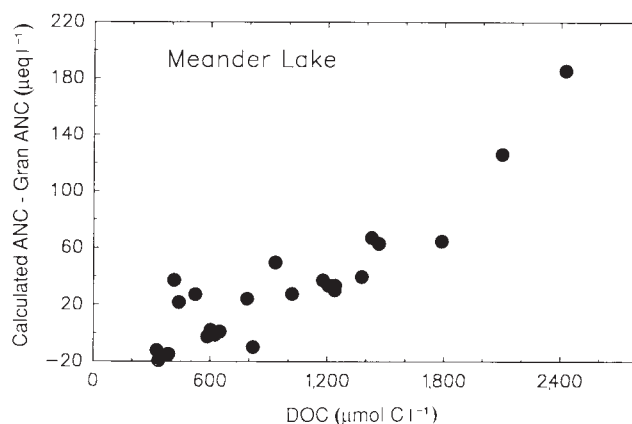


FIG. 3 The difference between calculated ANC (with 2 times Al_m) and Gran ANC versus DOC for waters from the Meander Lake basin.

concentrations of $0\text{--}25\ \mu\text{eq l}^{-1}$ ($\sim 0\text{--}300\ \mu\text{mol C l}^{-1}$). At higher organic-acid concentrations the difference increases, reaching 28% at $165\ \mu\text{eq l}^{-1}$ ($2,100\ \mu\text{mol C l}^{-1}$). However, when these same synthetic titration data were analysed using a standard laboratory Gran analysis algorithm, the differences were larger. For an organic-acid concentration of $25\ \mu\text{eq l}^{-1}$, the difference was 10%, and it increased to 44% at $165\ \mu\text{eq l}^{-1}$.

The increasing curvature of the F_1 function with increasing organic-acid concentration (Fig. 4b) can be problematic in analysis of titration data. The curve is nonlinear at much lower pH values (and thus higher F_1 and V values) as the DOC concentration increases. If the titration is not carried out to a

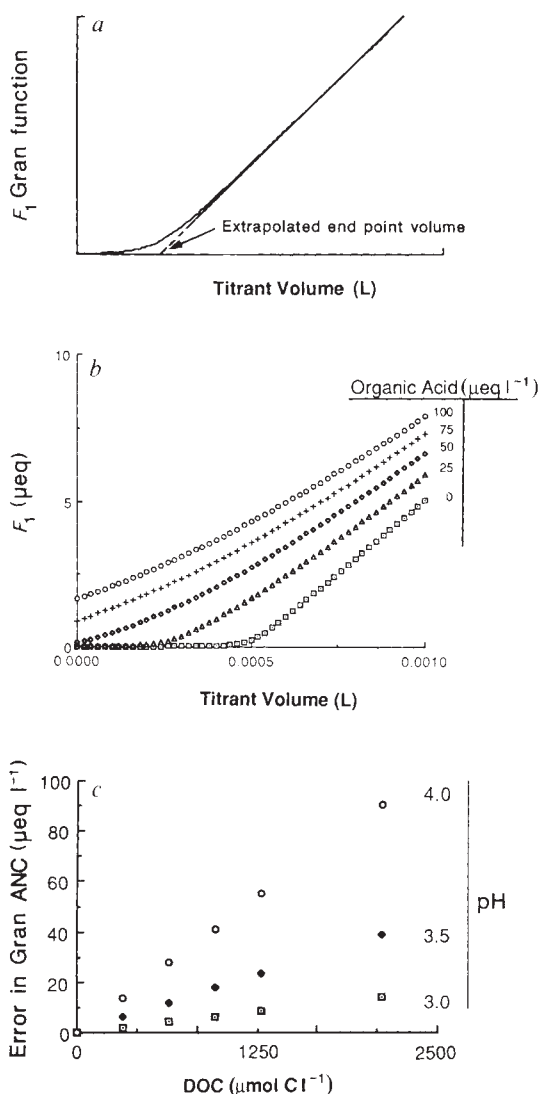


FIG. 4 Analysis of curvature in the Gran procedure: a, illustration of a Gran plot. The linear portion of the curve is extrapolated to determine the equivalent titrant volume. The slope of the line should correspond to the normality of the titrant. b, Curvature in F_1 Gran function due to organic acids. Curves are for carbonate solutions to which varying amounts of organic acid ($pK_A = 4.5$) have been added ($ANC = 50\ \mu\text{eq l}^{-1}$). At higher organic-acid concentrations, the curvature becomes more pronounced, and the intercept with the titrant volume axis shifts to lower values. c, Error in Gran ANC due to curvature of the F_1 function versus DOC, for different final titration pH values. Number of points in the F_1 Gran function used to extrapolate were 35 out of 75 (pH 4.0), 90 out of 190 (pH 3.5), and 300 out of 600 (pH 3.0). Failure to titrate to low pH values can lead to large errors in ANC.

TABLE 1 Gran analysis

Initial ANC ($\mu\text{eq l}^{-1}$)	Organic-acid concentration ($\mu\text{eq l}^{-1}$)	Organic-acid strength pK_A	Gran ANC ($\mu\text{eq l}^{-1}$)
50	0	—	50
50	25	6.5	50
50	50	6.5	50
50	25	5.5	49.8
50	50	5.5	49.5
50	25	4.5	47.9
50	50	4.5	45.7
50	75	4.5	43.6
50	100	4.5	41.5
50	165	4.5	36.0

low enough pH, the curvature will not be minimized and thus ANC will be underestimated (Fig. 4c). As indicated, the error in Gran ANC can exceed $90\ \mu\text{eq l}^{-1}$ for high DOC waters owing to curvature in the F_1 function alone, if titrations are only carried out to pH 4.0. The error decreases as the final pH of the titration decreases. Below pH 3.0, however, the error associated with the precise measurement of H^+ concentration becomes too large to further improve Gran estimates.

These calculations indicate that the curvature error can range from 0 to $90\ \mu\text{eq l}^{-1}$ for organic-acid concentrations of $0\text{--}165\ \mu\text{eq l}^{-1}$ ($0\text{--}2,100\ \mu\text{mol C l}^{-1}$). The actual differences observed in the Meander Lake basin waters were considerably larger. We hypothesize that the additional differences result from increasing strength of the acid functional groups as the solution H^+ concentration increases. This type of behaviour has been observed²⁶ and has been attributed to protonation of adjacent functional groups of polyprotic organic acids²⁷.

Received 25 August 1988; accepted 30 January 1989.

- Gran, G. *Analyst* **77**, 661–671 (1952).
- Omerik, S. M. & Powers, C. S. *Total Alkalinity of Surface Waters—National Map*, Corvallis Environmental Research Laboratory (US EPA, Corvallis, 1982).
- Goldstein, R. A., Gherini, S. A., Chen, C. W., Mok, L. & Hudson, R. J. M. *Phil. Trans. R. Soc. Lond.* **B305**, 409–425 (1984).
- Christophersen, N., Seip, H. M. & Wright, R. F. *Water Resour. Res.* **18**, 977–997 (1982).
- Reuss, J. O. & Johnson, D. W. *J. Environ. Qual.* **14**, 26–31 (1985).
- Cosby, B. J., Hornberger, G. M., Galloway, J. N. & Wright, R. F. *Water Resour. Res.* **21**, 51–63 (1985).
- Nikolaides, N. P., Rajaram, H., Schnoor, J. L. & Georgakakos, K. *Water Resour. Res.* **24**, 1983–1996 (1988).
- Smith, R. A., Alexander, R. B. & Wolman, M. G. *Science* **235**, 1607–1615 (1987).
- Charles, D. F. & Smol, J. P. *Limnol. Oceanogr.* **33**, 1451–1462 (1988).
- Sullivan, T. J. *Acid Deposition and Aquatic Ecosystems: Regional Case Studies* (ed. Charles, D. F.) (Springer, New York, in the press).
- Altshuler, A. P. & Linthurst, R. A. (eds) *The Acidic Deposition Phenomenon and its Effects: Critical Assessment Review Papers Vol. 2* (EPA, Washington, DC, 1984).
- Schindler, D. W. *Science* **239**, 149–157 (1988).
- Chen, C. W., Gherini, S. A., Munson, R. K., Gomez, L. & Donkers, C. *J. environ. Engng.* **114**, 1200–1216 (1988).
- Baker, J. P. & Schofield, C. L. *Water Air Soil Pollut.* **18**, 289–309 (1982).
- Brown, D. J. A. *Bull. Environ. Contam. Toxicol.* **30**, 582–583 (1983).
- Gherini, S. A. et al. *Water Air Soil Pollut.* **26**, 425–459 (1985).
- Stumm, W. & Morgan, J. J. *Aquatic Chemistry* (Wiley, New York, 1981).
- Church, M. R., Schofield, C. L., Galloway, J. N. & Cosby, B. J. *The Integrated Lake-Watershed Acidification Study*, Vol. 3, 7–1 to 7–7 (Electric Power Research Institute, Palo Alto, 1984).
- Schofield, C. L., Galloway, J. N. & Hendry, G. R. *Water Air Soil Pollut.* **26**, 403–423 (1985).
- Munson, R. K. & Gherini, S. A. *Acid Deposition and Aquatic Ecosystems: Regional Case Studies* (ed. Charles, D. F.) (Springer, New York, in the press).
- Goldstein, R. A., Gherini, S. A., Driscoll, C. T., April, R., Schofield, C. L. & Chen, C. W. *Biogeochemistry* **3**, 5–20 (1987).
- Driscoll, C. T. & Newton, R. M. *Environ. Sci. Technol.* **19**, 1018–1024 (1985).
- Schecher, W. D. & Driscoll, C. T. *Water Resour. Res.* **23**, 525–534 (1987).
- Driscoll, C. T. & Bisogni, J. J. *Modelling of Total Acid Precipitation* (ed. Schnoor, J. L.) 53–72 (Butterworth, Boston, 1984).
- Cook, R. B., Kelley, C. A., Kingston, J. C. & Kreis, R. G. Jr. *Biogeochemistry* **4**, 97–117 (1987).
- Oliver, B. G., Malcolm, R. L. & Thurman, E. M. *Geochim. cosmochim. Acta* **47**, 2031–2035 (1983).
- Stevenson, F. J. *Humus Chemistry* (Wiley, New York, 1982).

ACKNOWLEDGEMENTS. We thank M. Church, J. Eilers, R. Goldstein, G. Holdren, J. Butler, D. Landers and F. Unangst for assistance. The research described here was funded as part of the National Acid Precipitation Assessment Program (US Environmental Protection Agency through an interagency agreement with the US Department of Energy) and as part of the regionalized lake watershed acidification study (Electric Power Research Institute and the Empire State Electric Energy Research Corporation).

Unexpected changes in the oxic/anoxic interface in the Black Sea

J. W. Murray*, H. W. Jannasch†, S. Honjo‡, R. F. Anderson‡, W. S. Reeburgh§, Z. Top||, G. E. Friederich¶, L. A. Codispoti¶ & E. Izdar#

* School of Oceanography, University of Washington, Seattle, Washington 98195, USA

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York 10964, USA

§ Institute of Marine Sciences, University of Alaska, Fairbanks, Alaska 99701, USA

|| Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA

¶ Monterey Bay Aquarium Research Institute, Pacific Grove, California 93950-4616, USA

Dokuz Eylul University, Institute of Marine Sciences and Technology, 35213 Izmir, Turkey

THE Black Sea is the largest anoxic marine basin in the world today¹. Below the layer of oxygenated surface water, hydrogen sulphide builds up to concentrations as high as 425 μM in the deep water down to a maximum depth of 2,200 m (ref. 2). The hydrographic regime is characterized by low-salinity surface water of river origin overlying high-salinity deep water of Mediterranean origin^{1,3}. A steep pycnocline, centred at about 50 m is the primary physical barrier to mixing and is the origin of the stability of the anoxic (oxygen/hydrogen sulphide) interface. Here we report new observations, however, that indicate dramatic changes in the oceanographic characteristics of the anoxic interface of the Black Sea over decadal or shorter timescales. The anoxic, sulphide-containing interface has moved up in the water column since the last US cruises in 1969 and 1975. In addition, a suboxic zone overlays the sulphide-containing deep water. The expected overlap of oxygen and sulphide was not present. We believe that these observations result from horizontal mixing or flushing events that inject denser, saltier water into the relevant part of the water column. It is possible that man-made reduction in freshwater inflow into the Black Sea could cause these changes, although natural variability cannot be discounted.

The US-Turkish Oceanographic Expedition to the Black Sea consisted of five cruises on the RV *Knorr* from 15 April to 2 August 1988 (*Knorr* cruise 134, legs 8 to 12)⁴. One of the main goals was to study the anoxic interface in detail. The principal

features of the anoxic interface were summarized by Caspers¹. Oxygen was found to be depleted to zero concentration at depths of between 125 and 300 m. During the RV *Atlantis II* cruise in March and April 1969 the boundary between the oxygenated surface waters and the sulphide-containing deep waters was found to occur along a constant-density horizon. An overlap of oxygen and sulphide of a few metres was observed². The isopycnal surface (potential density, $\sigma_\theta = 16.41$) of the oxygen zero boundary was dome-shaped and ranged from a depth of 115 m in the centre of the basin to about 275 m near the shelves^{3,5}. The RV *Chain* visited the Black Sea in April 1975 and found hydrographic results that were nearly identical to those reported in 1969⁶. At that time the anoxic interface comprised a 20–30-m transition zone where oxygen and hydrogen sulphide coexisted at very low concentrations ($<15 \mu\text{M}$) (ref. 7). Russian scientists have frequently reported the coexistence of oxygen and hydrogen sulphide in a zone called the 'C layer' (see, for example, ref. 8).

Several studies in the early 1980s^{8–12} reported that the upper boundary of the sulphide zone was rising. We have confirmed such observations with high-quality continuous data and we find that the first appearance of sulphide in the water column is higher by as much as 30 m relative to the previous US studies in 1969 and 1975. In addition, we have also observed a significant gap of 10–40 m between the layer of oxygen decrease to suboxic levels (less than $5 \mu\text{M}$) and that of the first appearance of hydrogen sulphide. This gap was thicker in the centre of the Black Sea than near its boundaries.

During the 1988 RV *Knorr* cruises we studied the chemistry and microbiology of the interface zone in detail using a new pump-profiling system. This pump sampler was attached to a CTD (conductivity, temperature and depth instrument) and interfaced to an autoanalyser for detailed analyses of NO_3^- , NO_2^- , NH_4^+ , Si, PO_4^{3-} and H_2S . The pump was stopped at specific depths to collect samples for manual analyses. The pump was lowered at a rate of 6–10 m min^{-1} and data were acquired every 3 s, giving a sampling rate of 2–3 data points per metre. Under calm conditions, features in the water column could be resolved to better than 2 m. Within less than an hour the detailed structure of the water column was characterized on a down cast and a discrete sampling strategy designed for the subsequent up cast.

The potential temperature (θ) in the central part of the western basin of the Black Sea (station BS3-2; 32°E , $42^\circ50'\text{N}$; 5 June 1988)¹³ decreased from high surface values of $>17^\circ\text{C}$, reflecting seasonal warming, to a minimum of 6.95°C at 45 m, then increasing to 8.58°C at 100 m and to 8.9°C at 2,050 m (Fig. 1a). Station 1445 of the 1969 *Atlantis II* Black Sea cruise² ($31^\circ27'\text{E}$, $43^\circ08'\text{N}$; 1 April 1969) was in close proximity to the water BS3-2 and

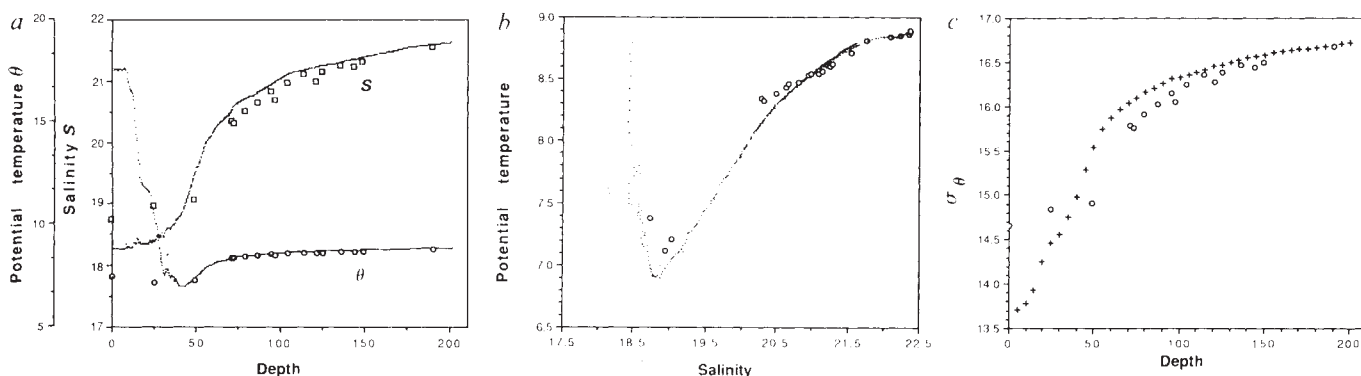


FIG. 1 Potential temperature, salinity and density in the Black Sea. Open symbols are from station 1445 of the March 1969 *Atlantis II* Black Sea cruise. Closed symbols are from station BS3-2 of cruise 3 of the US-Turkish expedition (June, 1988). a, Potential temperature (θ) and salinity (S) plotted

against water depth to 200 m, calculated from SeaBird CTD data with the assistance of G. White and M. Relander. b, Potential temperature against salinity for the entire water column. c, Potential density (σ_θ) against water depth to 200 m.

is used for comparison. The 1969 and 1988 data agree well in the range from 50 m and 100 m and below 200 m. From 100 m to 200 m the 1988 temperature is about 0.1 °C warmer. The surface temperatures are quite different because the cruises were made at different times of the year.

The salinity in 1988 increased from 18.33‰ at the surface to 21.09‰ at 100 m (Fig. 1a) and then to 22.33‰ at 2,050 m. The 1969 data are sparse above 70 m, but the salinity was definitely higher by at least 0.1‰ in 1988 over the 50–200 m depth range².

The change in hydrographic properties can also be seen by comparing plots of the potential temperature versus salinity (θ - S) (Fig. 1b). There is clearly a departure toward higher salinities at a given potential temperature for salinities greater than 21.00‰. The net result is that the potential density (σ_θ) was higher from 50 m to 200 m by up to 0.3 in 1988. This corresponds to an upward shift of the isopycnal surfaces in this depth interval by as much as 20 m. In 1969 the θ - S diagram was linear across the anoxic interface from 50 to ~400 m. In 1988 the θ - S data were linear from about 45 m to 70 m, and there was a small degree of curvature through the anoxic interface region. Since 1969 (and 1975) there has been a small but significant change in the hydrographic and density structure of the anoxic interface region (50–200 m) at this location in the Black Sea.

In comparison with the subtle changes in the water-column hydrographic and density structure, the changes in the oxygen, sulphide and nutrient distributions have been more pronounced. The new data show that dissolved oxygen is high in the surface water with a photosynthetically produced maximum at 10 m (Fig. 2a). Oxygen concentration decreases rapidly, to values of less than 5 μM at a depth of 55 m. In 1969 the measured oxygen concentration did not decrease to zero until about 125 m depth. Sulphide first appeared at about 95 m and increased almost linearly with depth (Fig. 2a). Repeat casts indicated that the uncertainty in this depth was about 5 m. This first appearance of sulphide occurred close to the density surface of $\sigma_\theta = 16.32$. In 1969 sulphide first appeared at 125 m, corresponding to the density surface of $\sigma_\theta = 16.39$. Thus sulphide appears to have shoaled slightly relative to the water-column density structure.

A prominent new feature in the anoxic interface is a suboxic zone that exists from 55 m to 95 m at station BS3-2. This zone is characterized by oxygen concentrations less than 5 μM , with no discernible gradients. This suboxic zone was found at all stations occupied during the 1988 expedition and was unexpected, because most recent literature^{3,14,15} reported the coexistence of O_2 and H_2S ; the vertical extent of this coexistence

was reported to be as large as 90–100 m (ref. 15) and appeared to vary laterally and seasonally⁸. There is one earlier report of a suboxic zone but it appeared to be limited to an isolated eddy experiencing unusually intense vertical mixing¹⁶.

Ammonia content (Fig. 2b) increased rapidly and linearly with depth below 80 m. The gradient relative to depth for ammonia in 1988 was similar to that in 1969², but was shifted upward by about 50 m in the recent data.

The 1988 phosphate profile (Fig. 2c) is intriguing because it shows a maximum at 55 m, decreasing to a minimum at 70 m, and then increasing rapidly to a large maximum at 95 m, coincident with the first appearance of sulphide. Similar profiles were found at most offshore stations. Examination of the 1969 data suggests that two PO_4^{3-} maxima were present but could not always be identified unambiguously because of the coarse sampling resolution¹⁴. The spline fit of the total 1969 data set used by Shaffer¹⁷ showed that they were probably present. The upper PO_4^{3-} maximum is probably due to release from decomposing organic matter during aerobic respiration¹⁴. The deeper PO_4^{3-} maximum probably results from reductive dissolution of solid Mn and Fe oxides, which release adsorbed PO_4^{3-} (ref. 17). Both PO_4^{3-} maxima appear to have risen in the water column by about 50 m relative to 1969.

The changes in the salinity and density structure may reflect horizontal mixing or flushing events that inject new water into the pycnocline. Several authors have proposed that the pycnocline is most probably influenced by lateral ventilation rather than by one-dimensional vertical transport (see, for example, ref. 18). The shallow temperature minimum bounded by the 8 °C isotherms is called the cold intermediate layer (CIL). At station BS3-2 the CIL is between 25 and 55 m. Tolmazin¹⁹ has summarized the evidence (for example, ref. 20) to suggest that the water in the CIL originates in the Northwestern Shelf (NWS) region of the Black Sea. Here, the coldest temperatures and low river runoff in winter result in the formation of dense water which descends the slope and spreads across the rest of the Black Sea on a timescale of about one year. Ovchinnikov and Popov²¹ have recently proposed the alternative view that the water in the CIL originates at the centres of the cyclonic gyres.

Regardless of the origin of the water in the CIL, its production rate and density must be closely linked to long-term variations in the freshwater and heat budgets of the Black Sea. When very dense water forms, the flow overshoots the traditional lower boundary of the CIL. Natural interannual or decadal variations in climate and river runoff can produce horizontal temperature and salinity (that is, density) waves in this shallow intermediate

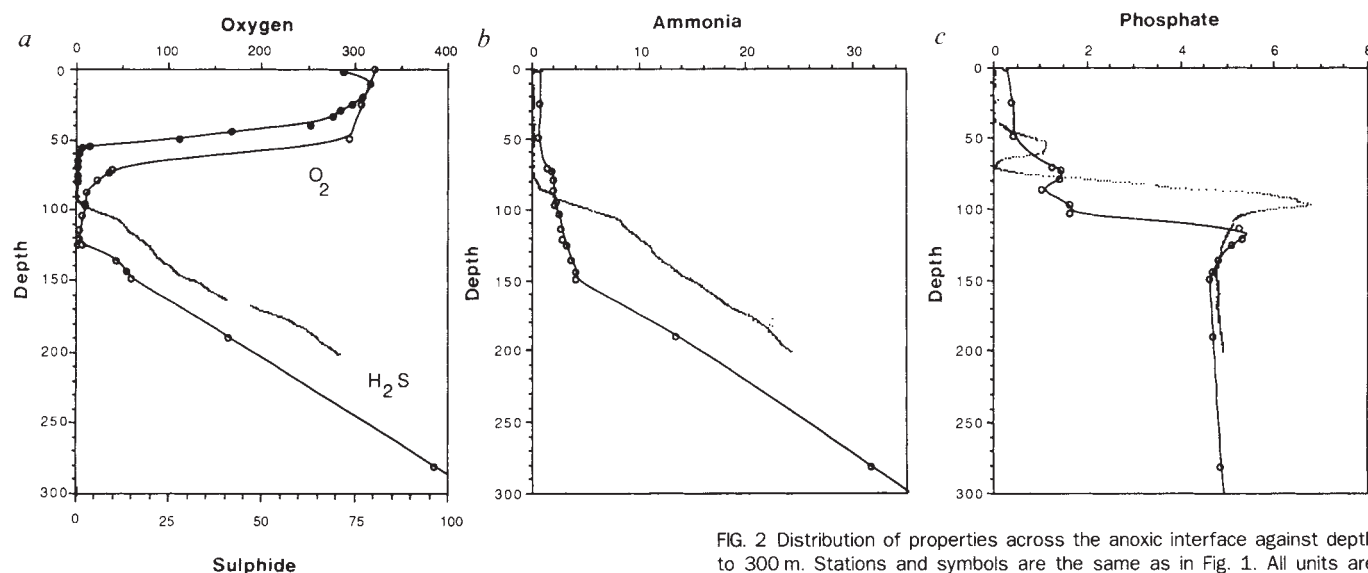


FIG. 2 Distribution of properties across the anoxic interface against depth to 300 m. Stations and symbols are the same as in Fig. 1. All units are $\mu\text{mol kg}^{-1}$. a, Oxygen and hydrogen sulphide. b, Ammonia. c, Phosphate.

layer. Superimposed on these natural variations is the man-made reduction in the freshwater runoff from Soviet rivers entering the NWS region of the Black Sea. Tolmazin^{19,22} reported that river flow of the Dnieper and Dniester had already been reduced by 50% in 1981. The total freshwater input from all rivers had been reduced by about 15%. Reduction in river input will also result in increased Mediterranean sea-water input through the Bosphorus²³. Substantial reduction in freshwater influx should cause changes in the salinity structure of the upper layer of the Black Sea. The extreme case of no river input would eventually result in a Mediterranean-style circulation for the Black Sea.

Our data clearly show that the salinity and density in the depth range 50–200 m was high in 1988 relative to historical values. It is possible that a reduction in the freshwater inflow is resulting in the formation of intermediate waters that have higher salinities and densities than in the past. The data set is not complete enough, however, to allow us to distinguish man-made effects from possible natural variations.

Increased production of water to the shallow intermediate layers could be viewed as analogous to the mixing or flushing events that occur in other anoxic basins, such as Saanich Inlet^{24–26} and the Baltic Sea²⁷. These flushing events produce suboxic zones similar to the one recently observed in the Black Sea. Flushing is thought to displace anoxic water upward in the water column. The sulphide in this displaced water then reacts with oxygen to form an oxygen-depleted layer²⁸. Any remaining oxygen is then consumed by aerobic respiration, thus producing a suboxic layer.

Additional chemical and microbiological results will need to be considered for a complete description of the changes in the

oxic/anoxic interface. The flushing-event hypothesis is attractive, however, because it can simultaneously explain the changes in salinity, density and concentration of oxygen, sulphide and nutrients. □

Received 1 December 1988; accepted 17 February 1989.

1. Caspers, H. *Geol. Soc. Am. Mem.* **67**, 803–890 (1957).
2. Brewer, P. G. *Woods Hole Oceanogr. Inst. Tech. Note* 71–65 (1971).
3. Spencer, D. W. & Brewer, P. G. *J. geophys. Res.* **76**, 5877–5892 (1971).
4. Murray, J. W. & Izdar, E. *Oceanogr. Mag.* (in the press).
5. Spencer, D. W., Brewer, P. G. & Sachs, P. L. *Geochim. cosmochim. Acta* **36**, 71–86 (1972).
6. Gagosian, R. B. & Heinzer, F. *Geochim. cosmochim. Acta* **43**, 471–486 (1979).
7. Karl, D. M. *Limnol. Oceanogr.* **23**, 936–949 (1975).
8. Fashchuk, D. Ya. & Ayzatulinn, T. A. *Oceanology* **26**, 171–178 (1986).
9. Boguslavskiy, S. G., Zhorov, V. A. & Novoselov, A. A. *Morsk. gidrofiz. zhurn* **1**, 54–58 (1985).
10. Novoselov, A. A., Sovga, Ye. Ye., Fashchuk, D. Ya., Khomutov, S. M. & Sheremet'yeva, A. I. *Ocnology* **27**, 304–307 (1987).
11. Leonov, A. V. & Ayzatulinn, T. I. *Oceanology* **27**, 174–178 (1987).
12. Zhorov V. A. *Geochim. Int.* 65–73 (1982).
13. Murray J. W. *School of Oceanography spec. Rep. No. 108* (Univ. of Washington, 1988).
14. Brewer, P. G. & Murray, J. W. *Deep Sea Res.* **20**, 803–818 (1973).
15. Vinogradov, M. Te., Shushkina, E. A., Flint, M. V. & Tumantsev, N. I. *Oceanology* **26**, 222–228 (1986).
16. Bol'shakov, V. S., Tolmazin, D. M. & Rozengurt, M. Sh. *Izvetia Acad. Sci. U.S.S.R. Geophys. Ser.* **6**, 562–565 (1964).
17. Shaffer, G. *Nature* **321**, 515–517 (1986).
18. Rooth C. G. H. in *Report on the Chemistry of Seawater. XXXIII* (Univ. of Goteborg, 1986).
19. Tolmazin, D. *Prog. Oceanogr.* **15**, 217–276 (1985).
20. Georgiev, Yu. S. in *Okeanograficheskiye issledovaniya Chernogo Morya* 105–113 (Naukova Dumka, Kiev, 1967).
21. Ovchinnikov, I. M. & Popov, Yu. I. *Oceanology* **27**, 555–560 (1987).
22. Tolmazin, D. *New Scientist* 767–769 (1979).
23. Tolmazin, D. *Prog. Oceanogr.* **15**, 277–316 (1985).
24. Emerson, S. et al. *Geochim. cosmochim. Acta* **46**, 1073–1079 (1982).
25. Tebo, B. M., Neilson, K. H., Emerson, S. & Jacobs, L. *Limnol. Oceanogr.* **29**, 1247–1258 (1984).
26. Anderson, J. J. & Devol, A. H. *Deep Sea Res.* **34**, 927–944 (1987).
27. Kremling, K. *Mar. Chem.* **13**, 87–108 (1983).
28. Anderson, J. J. & Devol, A. H. *Estuar. Coast. mar. Sci.* **1**, 1–10 (1973).

Static amorphization of anorthite at 300 K and comparison with diaplectic glass

Quentin Williams* & Raymond Jeanloz

Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

DIAPLECTIC glass, which is produced without fusion, is characteristic of meteorite impact sites in feldspar- and quartz-rich sites on the Earth and Moon; it has also been produced in laboratory shock experiments^{1–3}. Its presence in nature is thus considered to be strong evidence of exposure to a high-pressure shock wave generated by impact. The formation of such glasses has been universally interpreted in terms of the brief, high-temperature and high-pressure conditions characteristic of shock loading. Here we report that crystalline anorthite ($\text{CaAl}_2\text{Si}_2\text{O}_8$) becomes amorphous at static pressures of between 22 and 28 GPa at 300 K. The amorphization is associated with a pressure-induced increase in the coordination of silicon and aluminium by oxygen, from fourfold to five- and sixfold. The increase in coordination appears as the sample vitrifies at pressure; on decompression, the coordination reverts to fourfold. The vitrification is spatially heterogeneous, and may be crystallographically controlled. The formation of glass from crystalline material under static conditions at 300 K indicates that the high strain rates and temperatures associated with shock compression are not required for the creation of glasses without fusion.

The starting material used in this study is 'transitional' anorthite from Miyake-zima, Japan, and is identical to the sample material used in previous shock experiments⁴. Single crystals of this material, each about 15 μm in thickness, were

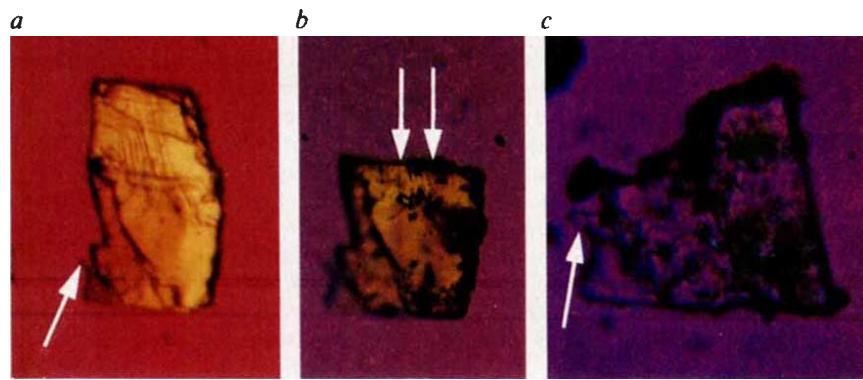
statically compressed in a Mao-Bell-type diamond cell; pressures were determined using ruby fluorescence⁵. A 16:3:1 mixture of methanol, ethanol and water was used as a pressure-transmitting medium, as this mix is hydrostatic to pressures of 15 GPa (ref. 6). At higher pressures, the samples were sufficiently small that the pressure gradients across them did not exceed 1 GPa. In particular, samples never bridged the gap between the diamond anvils. Pressures were increased over about 10 minutes, and the samples were then held at each pressure for approximately 18 hours, which is about 11 orders of magnitude longer than the corresponding periods in shock experiments. High-pressure infrared spectra were measured using a technique described elsewhere⁷, and indices of refraction were measured by the oil-immersion technique at ambient conditions.

Figure 1a shows a single crystal of anorthite before compression. The sample is optically anisotropic, with a birefringence of 0.013, in accord with the typical value for such crystals⁸. Samples compressed to 15 GPa exhibit no evidence for vitrification, nor any departure from single-crystal characteristics. After compression to 22 GPa, however, two optically isotropic bands appear in the sample (Fig. 1b). Each of these is about 20 μm wide, and they have a separation of about 40 μm . The birefringence of the crystalline material is indistinguishable (to within 20%) from that of the starting material. After compression to 28 GPa, the sample is optically isotropic, with a birefringence of less than 0.002 (Fig. 1c). Thus, crystalline anorthite is transformed to a glassy state merely by static compression to pressures exceeding 25 (± 3) GPa, at ambient temperature.

The index of refraction of a quenched isotropic sample of anorthite glass is 1.62 (± 0.02), significantly higher than the mean value of 1.58 observed for both anorthite crystal and shock-produced anorthite glass⁹. Thus, it seems that the amount of time over which a sample is compressed, and possibly the post-shock thermal regime, may profoundly affect the degree of densification within pressure-amorphized silicates. Comparisons between glasses formed by shocks of different duration might not, therefore, yield reliable estimates of peak shock pressures.

* Present Address: Institute of Tectonics, Board of Earth Sciences, University of California, Santa Cruz, California 95064, USA.

FIG. 1 Anorthite sample at zero pressure (outside the diamond cell) photographed with cross-polarized transmitted white light passing through a gypsum plate: *a*, before compression, *b*, after compression to 22 GPa, and *c*, after compression to 28 GPa. The sample is approximately 15 μm thick and each scale bar is 50 μm long. The multiply compressed sample fractured between photographs, but it is shown in the same orientation and arrows indicate the same feature in *a* and *c*. The two arrows in *b* indicate the vertically oriented isotropic bands that formed at high pressure. Note that the colour of the transmitted light is identical to the background colour in *c*, indicating that the sample is isotropic, except where small grains of ruby (used for pressure calibration) lie on its surface.



To determine the mechanism of the vitrification process, one must examine the microscopic structure of the sample as it amorphizes. Pressure-induced amorphization of two polymorphs of SiO_2 , quartz and coesite, has been documented by Raman spectroscopy and X-ray diffraction^{10,11}. As this amorphization involves the disappearance of both the X-ray diffraction pattern and, at higher pressures, the Raman spectrum, little information was obtained about the molecular-scale structure of the high-pressure amorphous phase. In contrast, infrared absorption spectra may characterize anorthite both before and after amorphization.

Figure 2 shows infrared spectra of initially crystalline anorthite as it is compressed to 1.7 GPa and 33.4 GPa, and then decompressed to 1.0 GPa. The strong absorption bands between 850 and 1,200 cm^{-1} in the low-pressure spectra of anorthite are due to antisymmetric stretching vibrations of both the SiO_4 and the AlO_4 tetrahedra¹². The intensity of these tetrahedral bands is decreased significantly in the spectrum of anorthite at 33.4 GPa, and there is an increased amplitude in the spectral region between 700 and 900 cm^{-1} . Although anorthite undergoes a symmetry-lowering phase transition at 2.7 GPa (ref. 13), the effect of this structural change on the vibrational spectrum is small (Q.W. and R.J., unpublished results). Absorption between 700 and 900 cm^{-1} is characteristic of silicon-oxygen and aluminium-oxygen polyhedra with octahedral and distorted octahedral coordination^{7,14}. At 33.4 GPa, the broadening of the bands and the appearance of new bands in the spectrum, combined with our optical microscopic studies, indicate that the sample has become amorphous at this pressure¹⁵. In the spectrum of decompressed anorthite, the tetrahedral peaks have returned to their initial amplitude but the entire spectrum is broader and more poorly resolved than that of the starting material. This demonstrates, in accord with our optical observations, that the quenched sample is amorphous. Thus, the initially fourfold-coordinated crystalline anorthite is transformed under pressure to a glass with an increased coordination of aluminium and silicon. This densified, highly coordinated glass subsequently reconverts to a tetrahedrally coordinated glass on decompression. Such reversible transitions from fourfold to sixfold coordination of silicon and aluminium have been observed at 300 K within a glass of anorthite composition at pressures above ~ 11 GPa (refs 7, 15).

Four mechanisms have been proposed previously for the creation of diaplectic glass: (1) reversion from a crystalline high-pressure phase on pressure release, (2) localized melting in 'shear bands', (3) defect-mediated amorphization, and (4) formation of a high-density glass under pressure. The first of these presumes that a crystalline high-pressure phase is formed under shock, which then reverts to glass during the high-temperature release from the shocked state. This model was proposed primarily because the density of framework silicates is large under shock compression, and it explicitly assumes thermodynamic equilibrium in the shocked state¹⁶. For com-

parison, the equilibrium crystalline assemblage of anorthite composition at pressures above 15 GPa is a three-phase mixture of CaSiO_3 -perovskite, corundum and stishovite¹⁷. The amorphization that we observe is completely distinct from a transition to these phases, however. Moreover, corundum and stishovite are quenchable down to 1 bar pressure, yet are not present in our decompressed samples. Thus, we conclude that the transition to an amorphous state occurs under pressure as a non-equilibrium, kinetically hindered process.

Formation of diaplectic glass by local melting in shear zones—regions of high temperature created by intense deformation at high strain rates—has been proposed on the basis that shock deformation is spatially heterogeneous^{9,18–21}. Both theoretical and empirical results support the shear-band model of shock deformation, but no correlation with glass formation has been demonstrated. Because neither high strain rates nor high temperatures are involved in glass formation in these experiments, we conclude that neither of the first two models, post-shock reversion of a high-pressure crystalline phase and localized melting in shear bands, satisfactorily explains the formation of pressure-induced glasses.

The last two mechanisms—defect-mediated processes (perhaps involving high dislocation densities) and direct conversion of low-pressure framework silicates to denser glasses under pressure—are both plausible means of amorphization^{3,22–24}. These mechanisms of glass formation are not mutually exclusive, and they are supported by our observations. For example, the infrared spectra indicate that the coordination of aluminium and silicon increases at high pressures; the more highly coordinated species must, when present in small concentrations, constitute defects in the feldspar structure. Neither of these mechan-

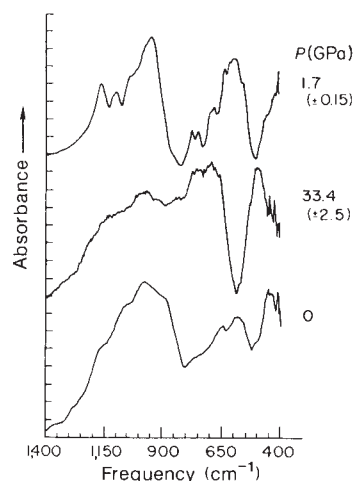


FIG. 2 Infrared absorption spectra of anorthite at 1.7 GPa (top), at 33.4 GPa (middle), and on subsequent decompression to 1.0 GPa (bottom), all at 300 K.

isms, however, predicts explicitly the spatially abrupt transition that we observe between crystal and glass²¹⁻²⁵. If defect generation (or increases in coordination) were homogeneous within the sample, amorphization would be expected to constitute a macroscopically uniform transformation. In this context, our observation that the amorphous regions are initially confined to sharp bands within the crystalline material affords an interesting comparison with shock-generated diaplectic glasses (Fig. 1b). Anorthite samples recovered from shock pressures of 30 GPa contain planar, crystallographically oriented lamellae of glass, juxtaposed with crystalline anorthite²⁵. Although the features observed in the shock-compressed samples are typically of sub-micrometre size, the planar features observed in our sample may be of similar origin. We speculate that, following the deformation of a tetrahedron or the collapse of a ring of tetrahedra to more highly coordinated species, subsequent deformations are controlled crystallographically and nucleate preferentially near the initial defect.

The mechanism proposed here by which crystalline feldspars are converted, via an amorphous, highly coordinated structure, to a glass provides a plausible explanation for the lack of significant argon loss from feldspars during the shock-loading process^{26,27}. Such loss has been reported when the sample melts under shock, as is expected in the shear-band model of diaplectic-glass generation, or at pressures above about 50 GPa, for which melting of feldspars is inferred to occur^{23,24}. In the formation mechanism proposed here, the coordination of the glass changes but there is little diffusional motion, and the transformation proceeds entirely in the solid state. Thus, we would expect shock-loading itself to induce little argon loss in feldspars at pressures under 50 GPa, although high post-shock temperatures may cause significant loss at lower pressures²⁶.

In the case of SiO₂, the properties of glassy and crystalline silica are such that the free energies of the ordered and disordered phases become similar at the pressure at which amorphization occurs. Thus, it has been proposed that the transition from high-pressure crystal to amorphous solid is described quantitatively by (metastable) equilibrium thermodynamics¹¹, with kinetic effects preventing formation of the stable crystalline phase(s). An analysis of the relative stabilities of anorthite and anorthite glass is consistent with our observations: using available elastic and thermodynamic properties^{13,28-30} and assuming a value of 4 for the derivative of the bulk modulus with respect to pressure for each phase, we predict a pressure of 28 GPa for the crystal-to-amorphous transition in CaAl₂Si₂O₈. Because the elastic properties of anorthite glass are poorly constrained²⁹, however, the agreement of this prediction with our observations may be fortuitous. Moreover, it is possible that non-hydrostatic stresses play a role in determining the pressure and temperature range, and perhaps even the occurrence, of this crystal-to-glass transition. We cannot preclude this possibility, but are unable to resolve such effects on the amorphization pressure. As the shocked state is also heterogeneous, with local regions of high differential strain¹⁹, we conclude that the different states of stress and strain rate in shocked and statically compressed samples are not significant in the formation of diaplectic glass.

The inferred coordination changes in these metastably compressed silicates probably involve significant density increases. Large increases in density are observed on the Hugoniot shock-compression curve at those pressures at which diaplectic glass is formed^{16,31}. We therefore attribute the observation of a densified, mixed-phase regime along the Hugoniot of silicates^{16,31,32} to the presence not of coexisting low- and high-pressure crystalline phases, but of a mixed-coordination glass created by shock-loading. This is supported by the extreme difficulty in quenching crystalline high-pressure phases from Hugoniot conditions^{33,34}.

The presence of diaplectic glass in natural samples has long been interpreted as field evidence for impact-induced shock-loading. Although our results show that diaplectic-like glass can

be formed statically, we emphasize that the high pressures and relatively low temperatures at which this transition takes place are unlikely to be brought about by any known tectonic or volcanic process, whether in subduction zones³⁵ or explosive eruptions³⁶. Thus it remains probable that such glasses are formed naturally only by impact-induced shock compression.

Received 28 September 1988; accepted 2 February 1989.

- DeCarli, P. S. & Jamieson, J. C. *J. chem. Phys.* **31**, 1675-1676 (1959).
- Milton, D. J. & DeCarli, P. S. *Science* **140**, 670-671 (1963).
- Stöffler, D. & Hornemann, U. *Meteoritics* **7**, 371-394 (1972).
- Jeanloz, R. & Ahrens, T. J. *Geophys. J. R. astr. Soc.* **62**, 529-549 (1980).
- Piermarini, G. J., Block, S., Barnett, J. D. & Forman, R. A. *J. appl. Phys.* **46**, 2774 (1975).
- Fujishiro, I., Piermarini, G. J., Block, S. & Munro, R. G. *VIII AIRAPT Conf. Proc.* (ed. Bergman, S.) **2**, 608 (1981).
- Williams, Q. & Jeanloz, R. *Science* **239**, 902-905 (1988).
- Tröger, W. E. *Optische Bestimmung der Gesteinsbildenden Minerale* (Schweizer, Stuttgart, 1971).
- Stöffler, D. *J. Non-Cryst. Solids* **67**, 465-502 (1984).
- Hemley, R. J. in *High Pressure Research in Mineral Physics* (eds Manghnani, M. H. & Syono, Y.) 347-359 (Am. geophys. Un., Washington, DC, 1987).
- Hemley, R. J., Jephcoat, A. P., Mao, H. K., Ming, L. C. & Manghnani, M. H. *Nature* **334**, 52-54 (1988).
- Iishi, K., Tomisaka, T., Kato, T. & Umegaki, Y. *Neues Jb. Miner. Abh.* **115**, 98-119 (1971).
- Angel, R. J., Hazen, R. M., McCormick, T. C., Prewitt, C. T. & Smyth, J. R. *Phys. Chem. Miner.* **15**, 313-318 (1988).
- Williams, Q., Jeanloz, R. & McMillan, P. J. *geophys. Res.* **92**, 8116-8128 (1987).
- Williams, Q. & Jeanloz, R. *Mater. Res. Soc. Final Prog. Abstr.* 312 (1987).
- Ahrens, T. J., Petersen, C. F. & Rosenberg, J. T. *J. geophys. Res.* **74**, 2727-2746 (1969).
- Liu, L. G. in *The Earth: Its Origin, Structure and Evolution* (ed. McElhinny, M. W.) 117-202 (Academic, Orlando, Florida, 1979).
- Grady, D. E. *J. geophys. Res.* **85**, 913-924 (1980).
- Grady, D. E., Murri, W. J. & DeCarli, P. S. *J. geophys. Res.* **80**, 4857-4861 (1975).
- Arndt, J., Hummel, W. & Gonzalez-Cabeza, I. *Phys. Chem. Miner.* **8**, 230-239 (1982).
- Diemann, E. & Arndt, J. *Phys. Chem. Miner.* **11**, 178-181 (1984).
- Ashworth, J. R. & Schneider, H. *Phys. Chem. Miner.* **11**, 241-249 (1985).
- Stöffler, D. *Fortschr. Miner.* **49**, 50-113 (1972).
- Stöffler, D. *Fortschr. Miner.* **51**, 252-289 (1974).
- Kitamura, M., Goto, T. & Syono, Y. *Contr. Miner. Petrol.* **61**, 299-304 (1977).
- Bogard, D., Hörz, F. & Stöffler, D. *Geochim. cosmochim. Acta* **52**, 2639-2649 (1988).
- Jessberger, E. K. & Ostertag, R. *Geochim. cosmochim. Acta* **46**, 1465-1471 (1982).
- Liebermann, R. C. & Ringwood, A. E. *Earth planet. Sci. Lett.* **31**, 69-74 (1976).
- Boslough, M., Rigden, S. M. & Ahrens, T. J. *Geophys. J. R. astr. Soc.* **84**, 455-473 (1986).
- Robie, R. A., Hemingway, B. S. & Fisher, J. R. *U.S. Geol. Surv. Bull.* 1452 (1979).
- Gibbons, R. V. & Ahrens, T. J. *Phys. Chem. Miner.* **1**, 95-107 (1977).
- Ahrens, T. J. *Science* **207**, 1035-1041 (1980).
- Kleeman, J. D. & Ahrens, T. J. *J. geophys. Res.* **78**, 5954-5960 (1973).
- Jeanloz, R. et al. *Science* **197**, 457-459 (1977).
- Schubert, G., Yuen, D. A. & Turcotte, D. L. *Geophys. J. R. astr. Soc.* **42**, 705-735 (1975).
- Wohletz, K. H. *Bull. volcan.* **48**, 245-264 (1986).

ACKNOWLEDGEMENTS. We thank R. Hemley and particularly D. Stöffler for helpful comments. This work was supported by the NSF.

Carbonate melt from the mantle in the volcanoes of south-east Zambia

D. K. Bailey*

Department of Geology, University of Reading, Whiteknights, Reading RG6 2AB, UK

RECENTLY Wallace and Green¹ reported the experimental formation of a dolomitic (Ca-Mg) carbonate melt in equilibrium with peridotite minerals in the range 21-30 kbar and 930-1,080 °C. These results confirm earlier deductions that, in the presence of CO₂, initial melts from the mantle would be carbonatitic^{2,3}, and extend the conditions of carbonate melt generation. The dolomitic melt composition also endorses earlier observations of quench dolomite after experimental melting of a natural garnet peridotite⁴. Dolomitic ashes, erupted from volcanoes near the confluence of the Rufunsa and Luangwa rivers in south-east Zambia^{5,6}, offer the nearest analogue to the experimental melt yet reported. Quenched melt droplets in the volcanics now reveal new evidence indicating a mantle source for this natural dolomite liquid. Specifically, I present here results which show that the liquid contains magnesiochromite crystals (52% Cr₂O₃) that match those in mantle peridotites, kimberlites and lamproites. In contrast with the experimental liquid, the natural dolomitic melt has a low iron content,

* Address from July 1989: Department of Geology, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK.

and high manganese and strontium, with alkalis virtually absent. High potassium activity is recorded, however, in the intensely metasomatized rocks around the main volcanoes. These differences suggest that the mantle source region chemistry differs from the high-sodium source envisaged in the experiments. The Zambian carbonatites thus reveal new aspects of carbonate melt and fluid activity in the Earth's mantle.

Volcanism in the Rufunsa region was approximately contemporaneous with the Chilwa Alkaline Province in Malawi (early Cretaceous, 135–105 Myr ago)⁷, and during the activity the three major centres were intruded by coarse-grained carbonatites, similar in all essentials to the classic intrusive complexes of Chilwa and elsewhere⁸. At Rufunsa, however, the intrusions were pierced by eruptions of pyroclastic dolomitic carbonatite unique to this locality. Sub-aerial volcanics of the same composition are preserved around the two southernmost complexes. Widely scattered smaller vents, ranging in size down to a few metres across, are filled solely with dolomitic tuffsite which is in most cases the product of a single eruptive episode. All the volcanological features indicate diatreme eruption through the crust by high-velocity, particle-charged streams of melt or fluid. Alkaline silicate magmatism (which normally accompanies carbonatite activity) is absent from the Rufunsa volcanism; this too is consistent with eruption directly from the site of melt formation. Activity was clearly localized by the intersection of the Luangwa and lower Zambezi rifts (Fig. 1). Rift intersections provide favourable sites for alkaline and carbonatite magmatism in continental interiors^{9,10}, where the complex stress field favours perforation of the crust by deep-seated magmatism. Both the volcanology and the tectonics therefore point to high-speed eruption of the Rufunsa carbonatite volcanics directly from the mantle^{5,6}. Eruption speeds for CO₂-rich (diatreme-forming) melts from the mantle have been calculated by different methods^{11–13}, and values of around 70 km h⁻¹ are “generally regarded as reasonable”¹². Solids would be entrained in a melt,

or fluid, or in melt + fluid, the appearance of fluid being strongly system-dependent¹⁴.

A striking example of the carbonatite tuffsite is preserved as one of the final eruptions through the axial zone of the Chasweta volcano. The rock is a mixture of angular accidental fragments (from the vent walls) and clear, rounded lapillae, in a dark red, fine-grained matrix (Fig. 2). Some lapillae have droplet shapes, and similar clear material also appears in intricate curved wisps and crescentic forms. Many lapillae enclose accidental fragments, typically of barite, coarse carbonate or feldspar. As with the other lapillae, the enclosing jackets consist mainly of very fine-grained colourless carbonate, with textures ranging from interlocking granular to trachytic. Deposition of carbonate melt on high-level accidental fragments is consistent with the proposal that separation of carbonate melt and fluid may occur only as the eruption nears the surface¹⁴. At this point, separated melt could be quenched almost instantaneously in supercooled fluid, which would explain the preservation of fragile melt wisps. The quenched melt is largely Mn- and Sr-rich dolomite (Table 1, column 3), whereas the surrounding matrix is iron-oxide-rich, implying that iron was strongly segregated into the transporting fluid phase. In most of the carbonate analyses, alkalis were either not detected or were present close to the detection limit.

Probably the most remarkable finding in the microprobe examination of the carbonate lapillae was the presence of chrome-rich spinel. The same mineral is found also in the intrusive tuffsites of Mwambuto, the major volcanic centre to the west of Chasweta. The spinel is dispersed as minute octahedra (<100 µm) through the carbonate; its distribution, euhedral form and the strong chemical zoning (Table 1, columns 4 and 5) indicate crystallization from the carbonate melt, probably on the liquidus. It has generally been accepted that ‘carbonatites are not chromium-rich’¹⁵; one published average¹⁶, for example, gives a Cr content of 66 p.p.m., compared with 117 p.p.m. for average igneous rock. (Rufunsa melt lapillae have an estimated Cr content of around 1,000 p.p.m.) The Rufunsa volcanics are the first recorded instance of magnesiochromite-bearing carbonatite melt. For carbonatites worldwide, the spinel always hitherto reported is magnetite¹⁷ (which has a very low Cr content). At Rufunsa, magnetite is common in the coarse-grained intrusive carbonatites, emphasizing that the magnesiochromite-bearing pyroclasts are not simply volcanoclastic equivalents of normal carbonatite but must have a different origin. Chrome-rich spinels of this composition are typically associated with ultramafic rocks of high-pressure origin, the compositions in Table 1 matching those reported from mantle xenoliths in kimberlites^{18,19}. The compositions found are the same as early-crystallizing spinels both in kimberlites them-

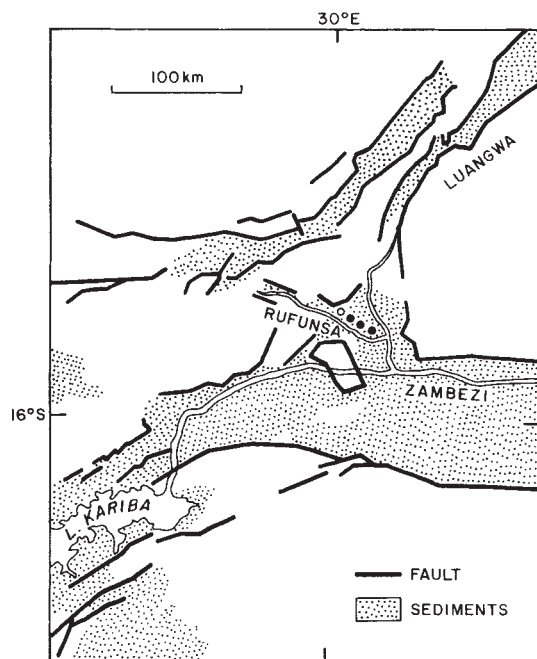


FIG. 1 Sketch map showing the intersection of the Luangwa (north-east-south-west) and Lower Zambezi Rift zones (rift depressions largely marked by sediments younger than 208 Myr (stipple), with older crystalline rocks (blank) forming higher ground): based on sheet 5 of ref. 29. The three main Rufunsa carbonatite volcanoes are indicated by filled circles. Sample H.780 comes from the vent of the most south-easterly volcano, Chasweta. The confluence of the Luangwa and Zambezi rivers marks the common border point of Zambia, Zimbabwe and Mozambique.

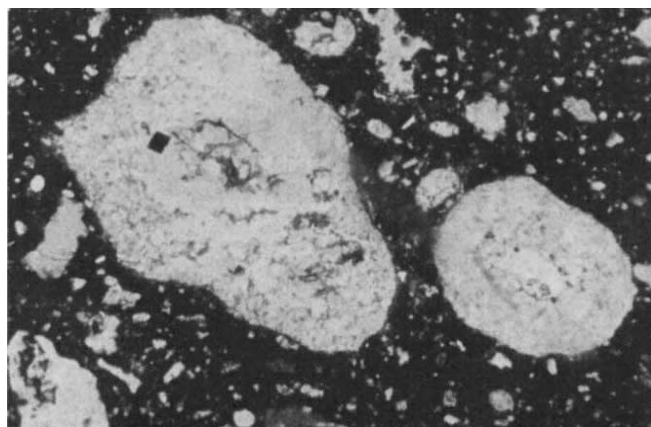


FIG. 2 Photograph of a thin section of the carbonatite volcanic in the vent of the Chasweta volcano. The rounded clear fragments are chilled carbonate melt droplets. Field of view is 3.4 mm wide.

selves²⁰ and in the 'group 1' spinels of lamproites²¹ (both magma types may be diamondiferous and hence of mantle origin). Two other characteristics of the Rufunsa spinels confirm their high-pressure source. Haggerty¹⁹ points out that low-pressure spinels are marked by higher levels of Ti and Fe³⁺, both of which are very low in the Rufunsa spinels. More importantly, he notes that spinels from the deepest mantle samples show a trend of falling Cr and rising Al with decreasing pressure and temperature, which is precisely the trend from core to rim in Rufunsa spinels. Rapid ascent of the Rufunsa melt, carrying the earliest-formed spinels, would provide such conditions of falling pressure and temperature, which are now recorded in the outer zones of the crystals by falling Cr and rising Al.

The association of carbonatites with mantle-derived magmas such as nephelinites, melilitites and kimberlites (which themselves may carry mantle xenoliths) puts their ultimate mantle source beyond question. In the Rufunsa province it is deduced, on the following grounds, that the carbonatite pyroclasts were erupted directly from the mantle source: (1) Rift intersection tectonics and diatremic style of eruption are consistent with direct tapping of the mantle^{5,6}; (2) The dolomitic composition is that predicted by experiment for carbonate melts from a high-pressure peridotite source^{1,4}; (3) The presence of magnesiochromite emphasizes the ultramafic connection; (4) Magnesiochromite compositions completely match those of high-pressure mantle rocks and accepted mantle melts (kimberlites/lamproites); (5) Zoning in the magnesiochromite is consistent with crystallization under conditions of falling *P* and *T* during magma ascent from the mantle source.

Attention has been drawn previously to the lack of sodium minerals in the Rufunsa volcanism^{6,22}, the only alkali silicates in these volcanics being the K-Mg mica, phlogopite, and pure potash feldspar⁶ (Table 1, columns 7 and 8). Both minerals are contained in fragments of metasomatized wall rocks. Phlogopite is very abundant in some eruptives, pointing to a greater abundance at depth and to the likelihood that the carbonatite melt source would have had phlogopite, rather than amphibole, as the major alkali-bearing phase. Experiments show that Mg-

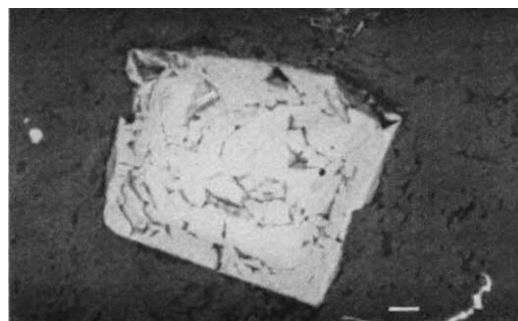


FIG. 3 Scanning electron microscope (back-scatter) image of a polished section of sample H.780, showing euhedral, zoned magnesiochromite in a carbonate matrix (analyses in Table 2). Scale bar, 10 μ m.

bearing carbonate melts must be expected from a phlogopite-bearing mantle^{23,24} at depths below the amphibole stability region¹, but analysis of the minute near-solidus liquids has not yet been possible.

For a more complete picture of the magmatic chemistry one must combine the melt composition, as revealed in the lapillae, with available data on the fluids in the system. In the dark red matrix of the volcanics, the iron content (as Fe₂O₃) may be as high as 10%, which must be ascribed largely to the transporting fluid, as no known country rock could provide an adequate source by contamination. Around the main vents, where there was sustained activity, potassium metasomatism affects all wall rocks regardless of initial composition, and its ultimate source must be in the magma system. Late fluids rich in Ba, P, S and F also cut all earlier rocks in the volcanoes. The total primary carbonatite budget should therefore include Mg, Ca, Mn, Sr, CO₂, P, Al, Cr and Fe in the melt, and K, Fe, Ba, S, P and F partitioned into fluids at lower pressure. Of these, only Mg and Cr (and possibly Al) would be expected in higher concentrations in the source mantle; all the other elements must be regarded as strongly partitioned from the source mantle into the carbonatite. Separation of melt and fluid at low pressures means that only the broad attributes of the total chemistry may be accessible from the volcanic system, but even this initial assessment reveals some remarkable parallels with melt inclusions in diamonds²⁵. Most notable in this respect is the balance of Ca and Mg, and the high levels of K, Fe, Ba and Sr in the diamond inclusions. Clearly, the Rufunsa carbonatites point the way to a new understanding of melt/fluid processes in the mantle.

There are many links between carbonatites and kimberlites, and in a few cases there are records of the presence of diamonds in the carbonatite association (refs 26, 27 and M. S. Garson, personal communication). It should be noted that kimberlites (of uncertain age) are also found along the Luangwa rift zone²⁸, some about 400 km to the south-west in Zimbabwe and some (diamondiferous) about 400 km to the north-east in Zambia. Typical indicator minerals of kimberlite, such as picro-ilmenite and pyrope, were not sought in the original survey of Rufunsa, when the chief economic targets were Nb and P mineralization. In view of the mantle origin of the carbonatite volcanics, however, the possibility cannot be excluded that the Rufunsa volcanics may themselves contain diamonds. □

TABLE 1 Carbonatite volcanic rock (H.780) and constituent minerals compared with experimental run products (*) at 22 kbar pressure and 1,000 °C

	1	2*	3	4	5	6*	7	8
SiO ₂	—	2.94	—	—	—	—	64.79	42.02
TiO ₂	—	0.45	—	1.29	1.78	0.61	—	0.97
Al ₂ O ₃	—	1.95	—	13.48	28.65	58.53	17.37	10.37
Cr ₂ O ₃	—	0.22	0.26	52.68	34.47	8.02	—	0.12
FeO	0.82	4.61	0.49	18.84	19.67	29.75	0.38	9.82
MnO	1.61	—	1.56	—	—	—	—	0.39
MgO	18.72	14.19	19.00	11.94	14.19	13.89	—	19.65
CaO	31.84	21.29	28.84	—	0.37	0.18	0.33	—
K ₂ O	—	0.35	—	—	—	—	16.40	10.63
Na ₂ O	—	4.99	—	—	—	—	tr.	0.34
P ₂ O ₅	0.45	0.48	—	—	—	—	—	—
SrO	—	—	1.10	—	—	—	—	—
BaO	—	—	—	—	—	—	0.32	0.22
Total	53.44	51.49	51.25	98.23	99.13	100.00	99.59	94.53

(1) CaO, MgO and MnO as carbonate in whole rock H.780 (ref. 5). Individual mineral analyses indicate small amounts of iron carbonate (compare column 3), and a mean value for FeO (~0.5 MnO) is applied here.

(2) Experimental carbonate melt in equilibrium with mantle silicates (Table 1, column 3 of ref. 1).

(3) Melt carbonate grain adjacent to chromian spinel (column 5), H.780.

(4) Magnesiochromite (core of crystal) in carbonate lapillus (Fig. 3).

(5) Chromian spinel (rim) in carbonate lapillus.

(6) Spinel in equilibrium with experimental carbonate melt (Table 1, column 9 of ref. 1). (Note that the spinel is aluminous, as expected for ~22 kbar pressure, in contrast with magnesiochromite typical of higher pressures¹⁹.)

(7) Potassium feldspar fragment in H.780.

(8) Phlogopite fragment in H.780.

Received 1 November 1988; accepted 17 February 1989.

- Wallace, E. W. & Green, D. H. *Nature* **335**, 343–346 (1988).
- Wyllie, P. J. & Huang, W. L. *Nature* **257**, 297–299 (1975).
- Eggler, D. H. *Geology* **4**, 787–788 (1976).
- Wendlandt, R. F. & Mysen, B. O. *Am. Miner.* **65**, 37–44 (1980).
- Bailey, D. K. *Bull. 5. Geol. Surv. Northern Rhodesia* 92 (1960).
- Bailey, D. K. in *Carbonatites* (eds Tuttle, O. F. & Gittins, J. 127–154 (Wiley, New York, 1966).
- Woolley, A. R. & Garson, M. S. in *African Magmatism and Tectonics* (eds Clifford, T. N. & Gass, I. G.) 237–262 (Oliver & Boyd, Edinburgh, 1970).
- Tuttle, O. F. & Gittins, J. (eds) *Carbonatites* (Wiley, New York, 1966).
- Bailey, D. K. *Geol. Mag.* **98**, 277–284 (1961).
- Bailey, D. K. *J. geol. Soc. Lond.* **133**, 103–106 (1977).

11. McGetchin, T. R. & Ullrich, G. W. *J. geophys. Res.* **78**, 1833–1853 (1973).
12. Mercier, J.-C. C. in *The Mantle Sample, Proc. 2nd Kimberlite Conf.* 197–212 (American Geophysical Union, 1979).
13. McCalister, R. H., Meyer, H. O. A. & Aragon, R. in *The Mantle Sample, Proc. 2nd Kimberlite Conf.* 244–248 (American Geophysical Union, 1979).
14. McGetchin, T. R., Nihhanj, Y. S. & Chodos, A. A. *J. geophys. Res.* **78**, 1854–1869 (1973).
15. Gittins, J. *Nature* **335**, 295–296 (1988).
16. Gold, D. P. *Miner. Soc. India, IMA Volume* 83–91 (1966).
17. Le Bas, M. J. in *Alkaline Igneous Rocks* (eds Fitton, J. G. & Upton, B. G. J.) 53–83 (Blackwell, Oxford, 1987).
18. Smith, J. V. & Dawson, J. B. *Phys. Chem. Earth* **9** (eds Ahrens, L. H., Dawson, J. B., Duncan, A. R. & Erlank, A. J.) 309–322 (Pergamon, Oxford, 1975).
19. Haggerty, S. E. in *The Mantle Sample, Proc. 2nd Kimberlite Conf.* 183–196 (American Geophysical Union, 1979).
20. Haggerty, S. E. *Phys. Chem. Earth* **9** (eds Ahrens, L. H., Dawson, J. B., Duncan, A. R. & Erlank, A. J.) 293–308 (Pergamon, Oxford, 1975).
21. Mitchell, R. H. *Trans. geol. Soc. S. Afr.* **88**, 411–437 (1985).
22. Bailey, D. K. *J. geol. Soc. Lond.* **145**, 103–105 (1988).
23. Wendlandt, R. F. & Eggler, D. H. *Am. J. Sci.* **280**, 421–458 (1980).
24. Olafsson, M. & Eggler, D. H. *Earth planet Sci. Lett.* **64**, 305–315 (1983).
25. Navon, O., Hutcheon, I. D., Rossman, G. R. & Wasserburg, G. J. *Nature* **335**, 784–789 (1988).
26. Griffin, W. L. & Kresten, P. in *Mantle Xenoliths* (ed. Nixon, P. H.) 101–106 (Wiley, New York, 1987).
27. Nixon, P. H. in *Mantle Xenoliths* (ed. Nixon, P. H.) 232 (Wiley, New York, 1987).
28. Nixon, P. H. in *Mantle Xenoliths* (ed. Nixon, P. H.) 187–193 (Wiley, New York, 1987).
29. *International Geological Map of Africa* (CGMW and UNESCO, 1986).

ACKNOWLEDGMENTS. I thank Iain Pryde for preparation of the fragile sample which made the initial microprobe analysis possible.

Mate guarding as paternity insurance in Idaho ground squirrels

Paul W. Sherman

Section of Neurobiology and Behavior, Seeley G. Mudd Hall,
Cornell University, Ithaca, New York 14853, USA

FOLLOWING a copulation, males in many species of vertebrates (particularly birds)^{1–4} and invertebrates^{5,6} remain near the inseminated female and repel other suitors with displays or force. Guarding males must delay resumption of competitive mate searching⁷, but they may insure their paternity by reducing possibilities for secondary matings and sperm competition^{8,9}. Among mammals, post-copulatory mate guarding has been reported in rodents^{10–12}, mongooses¹³, ungulates^{14,15} and primates^{16,17}, including humans¹⁸, but the effects of such behaviour on male reproductive success have not been determined genetically. I report here that mate guarding by male Idaho ground squirrels (*Spermophilus brunneus*) enhances a male's probability of paternity. Furthermore, an analysis based on game theory shows that mate guarding is an evolutionarily stable strategy for male *S. brunneus*, but not male Belding's ground squirrels (*S. beldingi*), which resume searching once copulation is completed.

The Idaho ground squirrel is the rarest and least known of the 18 North American *Spermophilus* species. It is endemic to an area of only ~8,000 km² in west-central Idaho, and the population is divided into small demes (many with <300 animals)¹⁹. During 1987 and 1988 I studied a deme of ~100 individuals located in a short-grass meadow at 1,870-m elevation (45°00' N, 116°40' W), 34 km northwest of Council, Idaho. Ground squirrels there are diurnally active for just over 4 months in the year, emerging from hibernation in late March and disappearing again by early August.

Every ground squirrel in the study population was marked with black hair dye for visual recognition and with eartags for permanent identification. The sexual behaviour of ground squirrels was observed on 13 days in 1987 (between 31 March and 14 April) and on 12 days in 1988 (4–17 April), for a total of 228 man-hours. Females were sexually attractive to males for only a few hours on the first or second afternoon following their emergence from hibernation. Newly emerged females remained near their hibernacula, where they were located and courted by adult males (≥2 years old), who themselves had emerged 1–2 weeks previously.

Copulations occurred below ground and so were not observable. Matings were inferred when four criteria were met²⁰: a male (1) followed a female closely and sniffed or licked her genitalia, then (2) accompanied her into a burrow, where (3) the pair remained for >5 min, after which (4) a sperm plug was observed in the female's vagina²¹. These criteria were fulfilled for each female on only one afternoon in the year, and pups invariably appeared above ground 50–52 days later. Thus I am confident that a female's oestrus and her copulatory behaviour were appropriately interpreted.

By watching continuously I recorded all male attendants to, and presumed copulations of, 12 females in 1987 and 14 females in 1988. Following matings, males stayed close to females: a male was ≤1 m from each focal female in 87% ± 15% (s.d.) of scan samples taken every 30–45 min after a female's first mating (total scans = 179). A given male attempted to keep his mate in a small area by herding, and he often followed her down a burrow then immediately re-emerged and blocked the entrance with his body. If a guarding male temporarily lost contact with his female (for example, if she went into a burrow unobserved), he gave a staccato chirping vocalization until she was relocated;

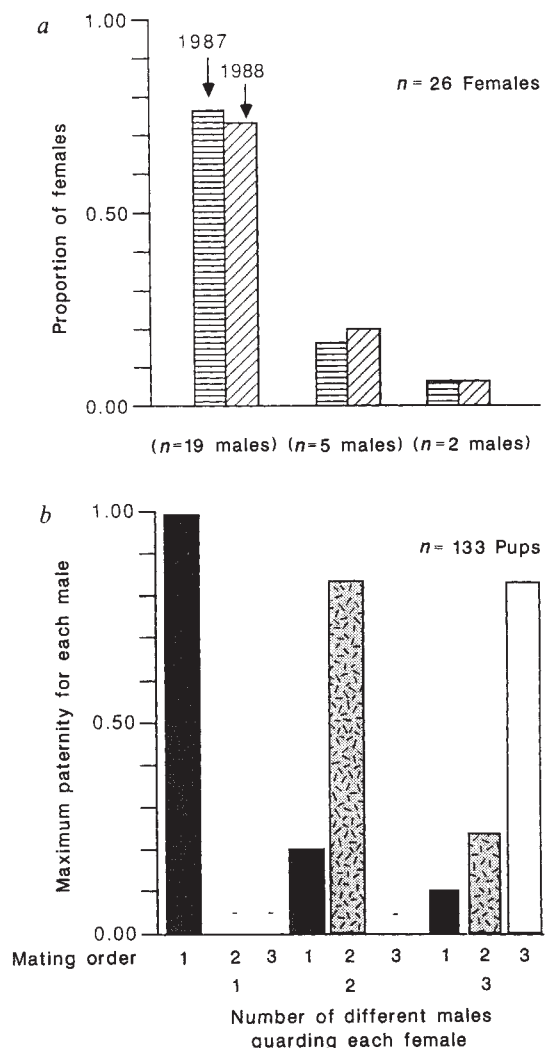


FIG. 1. a, The number of different males that guarded and apparently mated with 26 focal female Idaho ground squirrels (*S. brunneus*) in 1987 and 1988. b, Results of the electrophoretic paternity exclusion analyses, showing the highest proportion of the total offspring born to focal females that could have been sired by each successive guarding male (sample sizes are 97, 27 and 9 pups born to females guarded by 1, 2, and 3 males, respectively; data from 1987 and 1988 are combined).

the 'lost female' call was heard only in this context. Consort pairs went underground together every 39 ± 20 min. and spent 9.4 ± 1.3 min there. I captured 11 pairs the first time they resurfaced, and eight of the females had fresh vaginal plugs, indicating a recent copulation.

When a conspecific male approached a consort pair, the guarding male attacked him. In 72% of approaches (≤ 3 m; $n = 57$) the intruder fled, whereas in 28% he remained and fought. Usually males successfully sequestered their mate: of the 26 focal females, 19 were guarded by only one male and seven were guarded sequentially by multiple males (Fig. 1a). Guarding lasted 2.9 ± 1.1 h regardless of the number of males involved, and terminated only when females were no longer approached by rival males (that is, behavioural oestrus was over). During the 1987–88 mating periods the average number of females guarded by each male was 1.6 ± 0.9 , and the greatest number guarded successfully by any one male was four.

Successful guards had to repulse competitors 4.3 ± 1.7 times during their female's oestrus. Although males whose takeover attempts were thwarted did not sustain observable injuries in fights, it took them 1.3 ± 0.5 h ($n = 27$) to locate (come within 3 m of) another oestrus female, because females were scarce and scattered (Table 1).

Nine males began guarding females, but were displaced. Seven of these were ousted during the first third of the female's period of behavioural oestrus. Subsequent guards were always heavier than the male they displaced, and there was a positive correlation between the length of guarding and male body weight ($r = 0.79$, $P < 0.01$). Within 20 min after a change of the guard, the new male accompanied the female underground, where they presumably copulated. Ousted males immediately resumed searching, but every receptive female approached by these males was already being guarded, and only one was successful in taking over a female. Additionally, two were carried off by hawks (one by a prairie falcon, *Falco mexicanus*, and the other by a goshawk, *Accipiter gentilis*).

The day after a female was guarded she began constructing a nest burrow and excluding all conspecifics, both males and females, from a small territory surrounding it. Females reared their litter of 5.2 ± 1.4 young (range: 2–7, $n = 59$ litters) alone, and males did not live near their mates or behave parentally.

To investigate the genetic consequences of mate guarding, I live-trapped all 26 focal females, the 20 different males that guarded them, and all 133 pups they weaned (69 in 1987, 64 in 1988). A small sample of blood (~ 1 ml) was pipetted from the suborbital sinus of each animal, separated by centrifuging, and stored in liquid nitrogen (-196°C .) until analysis (8–10 weeks). Proteins were studied using standard horizontal starch-gel elec-

trophoresis and histochemical staining procedures^{22,23}. Of 32 putative loci assayed, seven were polymorphic in blood (allele frequencies are in parentheses): adenosine deaminase (0.64, 0.36), peptidase with glycyl-leucine (0.60, 0.40), peptidase with leucyl-glycyl-glycine (0.92, 0.08), esterase A (0.84, 0.16), protein 1 (0.85, 0.14, 0.01), protein 2 (0.79, 0.21), and protein 3 (0.92, 0.08). For these allozymes, frequencies of homozygotes and heterozygotes did not differ from Hardy-Weinberg expectation (all $P \geq 0.10$, χ^2 tests), and there was no evidence of linkage disequilibrium (all $P \geq 0.20$, G tests) or sex-linkage.

To analyse paternity, the phenotype of every animal was determined for all seven variable loci. Then, for each litter, the phenotype of every juvenile and its dam was used to implicate or eliminate each guarding male as the sire. The results (Fig. 1b) indicate that pups born to females that had been guarded by only one male were indeed sired by that male. Litters of females guarded by more than one male were typically multiply sired (five of the seven). However, paternal representation was nonrandom: a minimum of 66–100% of pups in each of these seven litters were sired by the last/longest attending male. These results could be due either to a preponderance of sperm from the male that guarded each female longest (that is, the one that apparently mated most often) or to last male sperm precedence^{24,25}. Preponderance is suggested by a nonsignificant but positive correlation between the number of (apparent) copulations by first guards and their maximum paternity among litters of multiply guarded females.

Male behaviour and paternity have been analysed in three other species of ground-dwelling sciurids. Black-tailed prairie dogs (*Cynomys ludovicianus*)^{26,27} and yellow-bellied marmots (*Marmota flaviventris*)^{28,29}, are harem polygynous: an adult male lives with a group of females throughout the year and protects them from intrusions by conspecific males and predators. The resident male sires the vast majority of the young and he behaves parentally toward them. Belding's ground squirrels (*S. beldingi*), by contrast, are a lekking species: males aggregate near females' hibernacula to mate, but subsequently do not live near those females or behave parentally^{30–32}.

In Belding's ground squirrels, receptive females were clustered (Table 1), and it took males only 0.4 ± 0.1 h ($n = 88$) after being defeated in a fight over a female to locate and initiate courtship with a new partner (significantly faster than male Idaho ground squirrels, $P < 0.001$, Mann-Whitney U test). Female *S. beldingi* typically mated with 3–5 partners, most litters were multiply sired³³, and paternity varied with mating precedence: a female's first mate sired on average 0.6 of the litter, her second mate 0.3, and her third and subsequent mates 0.1 (unpublished data). Male *S. beldingi* did not guard females, but

TABLE 1 Comparison of the availability and localizability of sexually receptive females in Idaho ground squirrels and Belding's ground squirrels

Species (study site)	Year (length of mating period)	Oestrus females available per day* (range)	Distance between burrows of receptive females† (m)	Average operational sex ratio‡ (range)
<i>S. brunneus</i> (Council, Idaho)	1987 (13 days)	1.4 ± 1.1 (0–4)	128 ± 39 ($N = 11$ female pairs)	1:3.2 (1:2–1:4)
	1988 (12 days)	1.6 ± 1.3 (0–5)	110 ± 34 ($N = 13$)	1:2.7 (1:2–1:4)
Significance level of comparison§		$\uparrow$ ($P < 0.01$)	$\uparrow$ ($P < 0.001$)	$\uparrow$ NS
<i>S. beldingi</i> (Tioga Pass, California ³⁰)	1975 (16 days)	2.6 ± 1.2 (1–6)	10 ± 3 ($N = 13$)	1:3.0 (1:2–1:6)
	1976 (15 days)	2.4 ± 0.8 (0–4)	7 ± 2 ($N = 9$)	1:3.2 (1:3–1:5)
	1978 (18 days)	2.2 ± 1.0 (1–5)	9 ± 1 ($N = 9$)	1:2.4 (1:2–1:4)

Males of the Idaho ground squirrel guard their mates after copulating, whereas males of Belding's ground squirrels do not. The study areas for both species were about half a hectare.

* Number of females that were guarded (*S. brunneus*) and mated (*S. beldingi*) on each day of the mating period.

† Nearest neighbour distances between pairs of females that were receptive on the same day.

‡ Mean daily ratio of sexually receptive females to adult males (≥ 2 yr old) during the mating period.

§ Significance levels for Mann-Whitney U test comparisons between the two species (data from different years combined).

resumed mate-finding and self-advertisement as soon as copulation was complete.

Using an approach based on game theory³⁴, I compared the effects of guarding versus searching on the fitnesses of male Idaho and Belding's ground squirrels (Table 2). These analyses suggest that mate guarding in Idaho ground squirrels is evolutionarily stable, largely because (1) locating receptive females is time-consuming for males and requires them to range widely (Table 1), which is dangerous, (2) obtaining sexual access to guarded females is difficult for all but the largest males, (3) unguarded females will mate multiply (Fig. 1a), thus diluting the first male's paternity, and (4) the longest (last) guarding male sires the majority of the litter (Fig. 1b). In contrast, conditions 1, 2, and 4 do not apply in Belding's ground squirrels and, as a result, resumption of mate searching after copulating is the evolutionarily stable state. The phenotypic costs and genetic consequences of mate guarding versus resuming searching merit investigation in other animals. □

Note added in proof: D. W. Foltz and P. L. Schwagmeyer have just reported³⁵ that $75 \pm 0.8\%$ ($n = 8$) of 13-lined ground squirrel

litters are sired by a female's first mate; this species thus closely resembles *S. beldingi* in both male post-copulatory behaviour and sperm usage pattern (Table 2).

Received 7 November 1988; accepted 15 February 1989.

1. Birkhead, T. R., Johnson, S. D. & Nettleship, D. N. *Anim. Behav.* **33**, 608–619 (1985).
2. Møller, A. P. *Behav. Ecol. Sociobiol.* **17**, 401–408 (1985).
3. Robinson, S. K. *Anim. Behav.* **34**, 241–255 (1986).
4. McKinney, F. in *Ecological Aspects of Social Evolution* (eds Rubenstein, D. I. & Wrangham, R. W.) 153–171 (Princeton Univ. Press, Princeton, 1986).
5. Waage, J. K. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 251–290 (Academic, Orlando, 1984).
6. Thornhill, R. & Alcock, J. *The Evolution of Insect Mating Systems* (Harvard Univ. Press, Cambridge, 1983).
7. Schwagmeyer, P. L. & Parker, G. A. *Anim. Behav.* **35**, 1015–1025 (1987).
8. Parker, G. A. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 1–60 (Academic, Orlando, 1984).
9. Birkhead, T. R., Pellat, J. & Hunter, F. M. *Nature* **334**, 60–62 (1988).
10. Farentinos, R. C. *Anim. Behav.* **20**, 316–326 (1972).
11. Barash, D. P. *Behav. Ecol. Sociobiol.* **9**, 187–193 (1981).
12. Dobson, F. S. *J. Mammal.* **64**, 218–225 (1983).
13. Rood, J. P. *Anim. Behav.* **28**, 143–150 (1980).
14. Lott, D. F. *Z. Tierpsychol.* **56**, 97–114 (1981).
15. Hogg, J. T. *Ethology* **75**, 119–144 (1987).
16. Tutin, C. E. G. *Behav. Ecol. Sociobiol.* **6**, 29–38 (1979).
17. Bercovitch, F. B. *Behav. Ecol. Sociobiol.* **21**, 163–172 (1987).
18. Flinn, M. V. *Ethol. Sociobiol.* **9**, 1–28 (1988).
19. Yensen, E. Unpubl. Rep. Idaho Dept. Fish & Game, 1–41 (1985).
20. Hoogland, J. L. & Foltz, D. W. *Behav. Ecol. Sociobiol.* **11**, 155–163 (1982).
21. Voss, R. *Occ. Pap. Mus. Zool. Univ. Mich.* **689**, 1–27 (1979).
22. Selander, R. K., Smith, M. H., Yang, S. H., Johnson, W. E. & Gentry, J. B. *Stud. Genetics, Univ. Texas Publ.* **7103**, 49–90 (1971).
23. May, B. P., Wright, J. E. & Stoneking, M. *J. Fish Res. Board Canada* **36**, 1114–1126 (1979).
24. Dewsbury, D. A. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 547–571 (Academic, Orlando, 1984).
25. Ginsberg, J. R. & Huck, U. W. *Trends Ecol. Evol.* **4**, 74–79 (1989).
26. Foltz, D. W. & Hoogland, J. L. *J. Mammal.* **62**, 706–712 (1981).
27. Loughry, W. J. *Behaviour* **103**, 27–48 (1987).
28. Schwartz, O. A. & Armitage, K. B. *Science* **207**, 665–667 (1980).
29. Armitage, K. B. in *Ecological Aspects of Social Evolution* (eds Rubenstein, D. I. & Wrangham, R. W.) 303–331 (Princeton Univ. Press, Princeton, 1986).
30. Sherman, P. W. thesis, Univ. Michigan (1976).
31. Sherman, P. W. & Morton, M. L. *Ecology* **65**, 1617–1628 (1984).
32. Sherman, P. W. in *Sociobiology: Beyond Nature/Nurture?* (eds Barlow, G. W. & Silverberg, J.) 505–544 (Westview, Boulder, 1980).
33. Hanken, J. & Sherman, P. W. *Science* **212**, 351–353 (1981).
34. Maynard Smith, J. *The Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
35. Foltz, D. W. & Schwagmeyer, P. L. *Am. Nat.* **133**, 257–265 (1989).
36. MacArthur, R. H. & Pianka, E. R. *Am. Nat.* **100**, 603–609 (1966).

ACKNOWLEDGEMENTS. I thank G. L. Bills, E. A. Domingue, J. E. and J. M. Dyer, J. L. Hoogland, C. Kagaire, Sherman, B. S. Mulder, D. M. O'Neill, H. K. Reeve, B. Semel, J. Shellman, Reeve, L. S. Sherman, M. S. Webster, D. F. Westneat, D. W. Winkler, E. Yensen and the owners of the OX Ranch. Electrophoretic analyses were performed in the Laboratory of Ecological and Evolutionary Genetics at Cornell, under the direction of B. P. May. I thank the National Geographic Society, the American Philosophical Society and G. C. Hixon for financial support.

Foraging specialization without relatedness or dominance among co-founding ant queens

Steven W. Rissing*, Gregory B. Pollock†, Mark R. Higgins*, Robert H. Hagen‡ & Deborah Roan Smith§

* Department of Zoology, Arizona State University, Tempe, Arizona 85287-1501, USA

† Department of Anthropology, Northwestern University, Evanston, Illinois 60208, USA

‡ Department of Entomology, Michigan State University, East Lansing, Michigan 48824, USA

§ Laboratory for Molecular Systematics, Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109, USA

HYPOTHESES on the evolution of sociality in the Hymenoptera have focused on two non-exclusive selective processes. First, individuals may help relatives to enhance inclusive fitness (kin selection^{1,2}). Second, group living may be so highly advantageous that competitively inferior individuals are forced into subordinate roles through social competition^{3–7}; in this hypothesis, subordinates help dominants in the expectation that they may benefit from the group's resources if the dominants lose status or die. Many

TABLE 2 An analysis of mate guarding behaviour versus resuming mate searching subsequent to copulating for male Idaho ground squirrels and Belding's ground squirrels, using game theory to find the evolutionarily stable state.

<i>S. brunneus</i>	<i>S. beldingi</i>
For guarding to be evolutionarily stable,	For not guarding (resuming searching) to be evolutionarily stable,
(1) $E(G, G) > E(N, G)$	$E(N, N) > E(G, N)$
Is it? If so,	Is it? If so,
(2) $[(g)(\Delta p)(l)]/t_g > [(1 - \Delta p)(l)]/t_s$	$[(1 - \Delta p)(l)]/t_s > [(g)(\Delta p)(l)]/t_g$
(3) $[(0.73)(0.83)(5.20)]/2.90 > [(0.43)(4.40)]/0.40$	$[(0.43)(4.40)]/0.40 > [(1)(0.57)(4.40)]/3.50$
(4) $1.1 > 0.7$	$4.7 > 0.7$
(units: offspring per hour)	(offspring per hour)
Conclusions	
Guarding is evolutionarily stable	Searching is evolutionarily stable

Schwagmeyer and Parker⁷ adapted the 'optimal diet' model of MacArthur and Pianka³⁶ to explore the relative payoffs of mate guarding versus resuming searching by male 13-lined ground squirrels (*S. tridecemlineatus*), a species in which males do not guard. They showed that guarding a female who has just been mated (G) is evolutionarily stable against invasion by non-guarding behaviour (N) when, in a population of guards, the expected reproductive benefits per unit time from guarding $E(G, G)$, exceed the expected gains per unit time from leaving the female and resuming searching $E(N, G)$, (inequality 1 for *S. brunneus*). Likewise, a non-guarding population (for example, *S. beldingi*) is resistant to the invasion of guarding behaviour when $E(N, N) > E(G, N)$. Expected gains from G and N behaviour depend on the parameters in inequality 2. For *S. brunneus*, these parameters are: g , the probability that a male can successfully guard his mate from other males, estimated as: $19/26 = 0.73$ (assuming that overthrown males gain no further matings); Δp , the increase in the probability of paternity due to guarding, calculated as: $1 - [\text{probability of paternity if the male does not guard}] = 1 - [(0.2) \times (\text{the likelihood that the abandoned female mates once more}) + 0.1 \times (\text{the likelihood that she mates twice more})] = 1 - [(0.2)(5/7) + (0.1)(2/7)] = 0.83$; see Fig. 1; l , the average litter size (5.2); t_g , the mean length of time a female is guarded (2.9 hours); and t_s , the mean time it takes a searching male to acquire a subsequent mate (estimated as 1.3 hours, the time it took defeated males to locate a new female). Note that both t_g and t_s are estimated conservatively relative to the inequality being examined; furthermore, predation risks, which make searching even more unfavourable, are not considered in the model. For *S. beldingi* the parameters in inequality 2 are defined similarly, but calculated differently: g is assumed to be 1.0 (that is, if males did guard females, they would do so effectively; this is the most conservative assumption relative to the inequality); Δp is calculated from the relative frequency of mating with 1–5 males³³ and the probability of paternity of the first male given that the female he leaves mates with 0–4 males subsequently (these values are $[(0.16)(1) + (0.26)(0.6) + (0.26)(0.3) + (0.26)(0.1) + (0.06)(0)] = 0.43$, so $\Delta p = 0.57$), l is 4.4 (ref. 31); t_g is assumed to be the length of the receptive period (3.5 h; ref. 30); and t_s is 0.4 hours.

social Hymenoptera associate predominantly with close relatives^{8,9}, which precludes an effective comparison of kin selection and social competition. Here we report on the existence of foraging specialists among unrelated co-foundresses of the leaf-cutter ant *Acromyrmex versicolor*; such task specialization leaves the forager at a relative fitness disadvantage within her foundress association. Contrary to the predictions of social competition theory, individuals specialize independently of competitive ability (as measured by relative body size) or reproductive status (as measured by ovarian condition) and without conflict. The selective basis of foraging specialization may lie in the intense competition that occurs among newly founded colonies engaged in reciprocal brood raiding.

In many ant species colonies are initiated by multiple foundresses¹⁰. As in other co-founding ants, *A. versicolor* foundresses do not associate preferentially by collection locale in laboratory experiments, suggesting that associations need not consist of close relatives^{11,12}. To confirm this, we estimated within-group relatedness¹³ from 26 field-caught foundress associations using allozymes as genetic markers. Of 30 tested, one locus was polymorphic¹⁴. Co-foundresses were no more closely related than randomly selected queens ($\hat{r} = -0.12$, s.e. = 0.03, estimated by a jack-knife over groups¹⁵). The negative $\hat{r}$ and small s.e. imply that genotypes are distributed more evenly across groups than would be expected under random association. The negative $\hat{r}$ is probably a sampling artefact caused by small group size ($N = 3.8$) and unequal allele frequencies; we do not infer that kin avoid each other¹⁶. We conclude that associations constitute random samples of the population, precluding kin selection as an explanation for social traits¹⁷⁻¹⁹.

Among co-founding ants, *A. versicolor* is unique in that foundresses regularly leave the nest to forage before the emergence of the first workers. Foragers are exposed to greater risks of predation and parasitization, and to greater thermal and physical stresses than their sedentary nest-mates²⁰⁻²². Nonetheless, leaves procured through foraging are shared communally in a single fungus garden. Foraging thus enhances colony fitness at the expense of a forager's relative intracolony fitness^{18,19}. As among co-founding wasps, foraging is a task in which specialization might indicate an incipient reproductive division of labour^{5,23}, a likely precursor to the evolution of permanent worker sterility which is characteristic of most social hymenoptera^{5,20,23,24}.

To study the distribution of foraging effort within colonies, we established eleven colonies, each with three newly mated foundresses collected immediately after a mating flight on 23 September 1987 at North Scottsdale, Arizona. To ensure substantial size variation within each association, the mass of each foundress in an association was arranged to be at least ± 1 standard deviation (equal to 1.04 mg) from the mass of the other two foundresses. Foundresses were marked, housed in 'ant

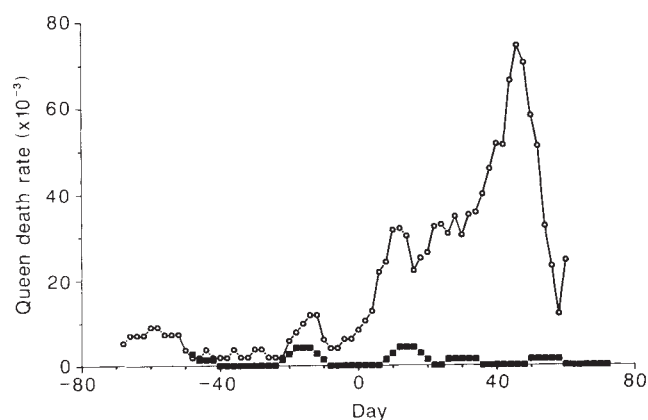


FIG. 1 Comparative rates of foundress mortality in *A. versicolor* (squares) and the ant *Veromessor pergandei* (circles), which displays more typical behaviour for cooperatively founding ants (data for *V. pergandei* from ref. 25). Queen death rate is calculated as the percentage of all the foundresses alive at the start of a 2-day period who died during that period; it is displayed as a 3-point moving average. All observations in each nest are standardized to day of first worker emergence (day 0) in that nest. *V. pergandei* and most other cooperatively founding species reduce to secondary monogyny soon after first worker emergence²⁸.

farms' and offered fresh native vegetation in attached foraging arenas, according to standard procedures²⁵. Farms and associated arenas were placed on a mechanized circular surface which rotated under a video camera. Foraging trips taken by each co-foundress in each association were censused from video records of 6-hour foraging sessions (57 sessions over 14 weeks). Sessions began immediately after colony formation and ended when workers assumed foraging tasks. Farms were not offered food between sessions. In each association virtually all foraging trips (average was 165 ± 69 trips per farm) were performed by a single foundress ($P < 0.001$ in each case); the only exceptions occurred in two associations where a single secondary forager appeared after the death of the first foraging specialist.

Another 34 farms were similarly established to see if specialization in foraging was a function of potential competitive ability, as measured by relative size. These farms were spot-sampled 8 times per day for 15 weeks to estimate the number of foraging trips taken by each foundress. Chi-square analyses indicated the existence of a foraging specialist in 44 of the 45 farms in the two samples combined; the role of foraging specialist was independent of initial relative size (Table 1). At the conclusion of these experiments, all co-foundresses from 18 associations were dissected to measure primary oocyte number and size

TABLE 1 Social correlates of reproductive status among *A. versicolor* co-foundresses

	Relative initial mass*			Foraging status	
	Large	Medium	Small	Forager	Non-foragers
Forager	18	16 ($\chi^2 = 2.36$; NS)	10		
Number of primary oocytes ($\pm$ s.d.; N)	8.2 (2.9; 18)	8.6 (1.9; 18) ($F = 0.01$; NS)	8.4 (2.1; 18)	8.6 (2.2; 18)	8.5 (1.9; 35)
Primary oocyte length, mm ($\pm$ s.d.; N of queens)	0.51 (0.11; 16)	0.53 (0.12; 16) ($F = 0.70$; NS)	0.56 (0.17; 16)	0.52 (0.10; 16)	0.54 (0.12; 32) ($t = 0.38$; NS)
Number of total oocytes ($\pm$ s.d.; N of queens)	18.56 (13.32; 18)	20.28 (16.43; 18) ($F = 0.06$; NS)	19.50 (14.56; 18)	20.37 (15.34; 19)	18.94 (14.32; 35) ($t = 0.34$; NS)

* Each foundress in each association was ≥ 1 s.d. in initial mass from the other 2 foundresses in that association, such that the total number of 'large', 'medium' and 'small' foundresses was the same over all foundress associations.

(routine assays of reproductive conditions^{20,26,27}). Reproductive condition was independent of body size and foraging specialization (Table 1); all foundresses had well developed ovaries and therefore presumably functioned as queens. During spot sampling, farms were examined for evidence of aggression that might induce a 'loser' to become a foraging specialist; in addition, the interiors of two colonies were videotaped for an average of 79 h per colony. We never saw aggression among foundresses, nor acts of ritualized dominance such as occur among some Polistine wasps^{5,23,27} (for example, asymmetric antennation, allogrooming or trophallaxis). Access to brood or fungus, the principal resources of a colony, was never contested by, nor denied to, any co-foundress. Unlike other cooperatively founding ants^{25,28} and some bees²⁰ and wasps²⁹, *A. versicolor* foundresses do not fight on first worker emergence (Fig. 1), and they share resources cooperatively for at least 1.5 yr in laboratory colonies.

The primary benefit of Hymenopteran foundress associations is believed to be defence against enemies^{20,27,30}; in some ants, the enemy is conspecific. Newly founded nests are often clumped spatially, and they engage in reciprocal brood raids; colonies with more workers (and queens to produce them) tend to prevail^{10,25,31}. This pattern is seen clearly in *A. versicolor*, where newly founded colonies are highly clumped¹², yet adult colonies are territorial³²; from a group of new colonies, only a single adult colony will survive. We marked workers in 16 ant farms and paired them to common foraging areas. Brood, leaf and fungus raids occurred. Of 163 raids, 131 involved marked workers; these workers always displayed fidelity to their natal colony. Fighting among workers, and between workers and foreign queens, was common. In each pairing only a single colony survived.

As only one colony from an initial cluster of colonies will survive to adulthood, relatively small differences in colony efficiency could cause large differences in ultimate colony reproductive success. An increase in foraging efficiency should translate to a similar increase in quality of initial raiding force and, ultimately, success in eliminating rivals. Foraging social insects enhance their efficiency through experience^{26,33,34}. An *A. versicolor* foraging queen gains experience as she gathers leaves; to refuse further foraging squanders the former resource. Yet each queen should prefer her colony mates to forage, but the simultaneous expression of such preferences would reduce colony fitness; energy and time spent in intragroup conflict over task allocation would limit the colony's future competitive potential^{11,35}. Under these circumstances, a foundress may be relatively indifferent as to whether she is the one that suffers the mortality risk associated with foraging, because the colony will either succeed as a unit or (more probably) fail as a unit. □

Received 14 October 1988; accepted 13 February 1989.

- Hamilton, W. D. *A. Rev. Ecol. Syst.* **3**, 193–232 (1972).
- West Eberhard, M. J. *Q. Rev. Biol.* **50**, 1–33 (1975).
- West Eberhard, M. J. *Science* **200**, 441–443 (1978).
- West Eberhard, M. J. *Proc. Am. phil. Soc.* **123**, 222–234 (1979).
- West Eberhard, M. J. in *Natural Selection and Social Behavior* (eds Alexander, R. D. & Tinkle, D. W.) 3–17 (Chiron Press, New York, 1981).
- Vehrencamp, S. L. *Anim. Behav.* **31**, 667–682 (1983).
- Fletcher, D. J. C. & Ross, K. G. *A. Rev. Entomol.* **30**, 319–343 (1985).
- Gamboa, G. J., Reeve, H. K. & Pfennig, D. W. *A. Rev. Entomol.* **31**, 431–454 (1986).
- Breed, M. D. & Bennett, B. in *Kin Recognition in Animals* (eds Fletcher, D. J. C. & Michener, C. D.) 243–285 (Wiley, Chichester, 1987).
- Rissing, S. W. & Pollock, G. B. in *Interindividual Behavioral Variability in Social Insects* (ed. Jeanne, R. L.) 179–222 (Westview, Boulder, 1988).
- Rissing, S. W. & Pollock, G. B. *Anim. Behav.* **34**, 226–233 (1986).
- Rissing, S. W., Johnson, R. A. & Pollock, G. B. *Psyche* **93**, 177–186 (1986).
- Pamilo, P. *Genetics* **107**, 307–320 (1984).
- Hagen, R. H., Smith, D. R. & Rissing, S. W. *Psyche* (in the press).
- Crozier, R. H., Pamilo, P. & Crozier, Y. *Behav. Ecol. Sociobiol.* **15**, 143–150 (1984).
- Wilkinson, G. S. & McCracken, G. F. *Evolution* **39**, 1169–1174 (1985).
- Wilson, D. S. *Am. Nat.* **111**, 157–185 (1977).
- Wilson, D. S. *A. Rev. Ecol. Syst.* **14**, 159–187 (1983).
- Wade, M. J. *Am. Nat.* **125**, 61–63 (1985).
- Michener, C. D. *The Social Behavior of the Bees* (Belknap Press of Harvard University Press, 1974).
- Gamboa, G. J., Heacock, B. D. & Wiltjer, S. L. *J. Kans. Entomol. Soc.* **51**, 343–352 (1978).
- Traniello, J. F. A., Fujita, M. S. & Bowen, R. V. *Behav. Ecol. Sociobiol.* **15**, 65–68 (1984).
- West Eberhard, M. J. *J. Kans. Entomol. Soc.* **51**, 832–856 (1978).

- Michener, C. D. in *Experimental Behavioral Ecology and Sociobiology* (eds Hölldobler, B. & Lindauer, M.) 293–305 (Sinauer, Sunderland, 1985).
- Rissing, S. W. & Pollock, G. B. *Anim. Behav.* **35**, 975–981 (1987).
- Jeanne, R. L. *A. Rev. Entomol.* **25**, 371–396 (1980).
- Itô, Y. in *Animal Societies: Theories and Facts* (eds Itô, Y., Brown, J. L. & Kikkawa, J.) 17–34 (Japan Sci. Soc., 1987).
- Hölldobler, B. & Wilson, E. O. *Naturwissenschaften* **64**, 8–15 (1977).
- Pfennig, D. W. & Klahn, J. E. *Z. Tierpsychol.* **67**, 198–203 (1985).
- Weislo, W. T. *Behav. Ecol. Sociobiol.* **15**, 157–160 (1984).
- Pollock, G. B. & Rissing, S. W. *Am. Nat.* **133**, 61–70 (1989).
- Gamboa, G. J. thesis, Arizona State Univ., 1974.
- Menzel, R. in *Experimental Behavioral Ecology and Sociobiology* (eds Hölldobler, B. & Lindauer, M.) 55–74 (Sinauer, Sunderland, 1985).
- Traniello, J. F. A. in *Interindividual Behavioral Variability in Social Insects* (ed. Jeanne, R. L.) 91–112 (Westview, Boulder, 1988).
- Pollock, G. B. & Rissing, S. W. in *The Ecology of Social Behavior* (ed. Slobodchinkoff, C. N.) 315–334 (Academic, San Diego, 1988).

ACKNOWLEDGEMENTS. J. M. Scriber provided electrophoresis facilities. G.B.P. was supported by a MacArthur Fellowship in International Peace and Security studies granted by the Social Science Research Council. S.W.R. was supported by the NSF.

Glycine enhances NMDA-receptor mediated synaptic potentials in neocortical slices

Alex M. Thomson, V. E. Walker* & Diana M. Flynn*

Department of Physiology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, UK

* Department of Physiology, University of Wales College of Cardiff, P.O. Box 902, Cardiff CF1, 1SS, UK

ONE class of excitatory amino-acid receptors, the *N*-methyl-D-aspartate (NMDA) receptors, mediates transmission at a small, but important, group of synapses in the neocortex^{1–3}. These receptors are implicated in neuronal plasticity during development in young mammals and in memory acquisition in adults^{4–6}. Recently, responses of isolated membrane patches to NMDA were shown to be greatly enhanced by glycine⁷. This, together with the demonstration that the strychnine-insensitive glycine-binding site is distinct from, but linked to, the NMDA receptor⁸ has excited intense interest in glycine as a synaptic modulator. Before proposing a physiological function, however, it is important to determine whether glycine could enhance synaptic responses to NMDA receptor activation in intact, adult tissue. An earlier study⁹ failed to demonstrate enhancement of NMDA responses when glycine was applied and it was proposed that in intact tissue the high-affinity glycine site was already saturated by endogenous glycine. It remained possible that glycine concentrations can be maintained at low levels close to synaptic receptors. We have examined responses of neurons in slices of adult neocortex to focal applications of excitatory amino acids and glycine and report enhancement by glycine of NMDA receptor-mediated excitatory postsynaptic potentials.

So that small synaptic responses could be recorded intracellularly under controlled conditions, we used slices of adult neocortex, which does not contain any strychnine-sensitive glycine-binding sites¹¹. The perfusing fluid was delivered at a low flow rate (less than 0.3 ml min⁻¹) to allow metabolic products, transmitters and modulators to be controlled by natural release, degradation and uptake processes, rather than the perfusion system.

In all cells tested in which rapid-onset, repeatable responses to electrophoretically applied excitatory amino acids could be evoked, glycine enhanced responses to NMDA (mean increase in amplitude 100.00 ± 65.75% s.d., range 22–284%, *n* = 38), but produced no significant change in responses to quisqualate (mean increase 4 ± 28.54%, range, 50% decrease to 20% increase). The effective glycine concentration is not known, but maximal effects were produced with ejection currents of less than 4 nA (Fig. 1). The responses when glycine was dissolved in acidic and alkaline solutions were similar. The effects of

glycine on NMDA responses were reversed readily when application was terminated. They were rapidly and reversibly blocked by kynurenate, an endogenous antagonist at the high-affinity glycine receptor⁹⁻¹⁰, applied electrophoretically ($n=29$) or in the bathing medium ($n=3$) (50–100 μ M, Fig. 1), but were unaffected by strychnine. Effective glycine currents produced no measurable change in resting potential or input resistance.

Stimulation of the white matter evoked excitatory postsynaptic potentials (e.p.s.ps) of between 1 and 5 mV. Glycine applied with a current that augmented responses to NMDA but not to quisqualate augmented the e.p.s.ps evoked in 5 of 24 of these cells (see Fig. 2a–e) and depressed the e.p.s.ps in 7 cells, without affecting the response of the cell to injected current. In a further 5 cells, larger applications of glycine (5 to 40 nA ejection current) caused decreases in both e.p.s.p. and current-pulse response (Fig. 2i, j). Glycine produced no change in the e.p.s.ps in the remaining 7 cells. Effects on e.p.s.ps were readily reversible.

All e.p.s.ps that were insensitive to, or depressed by, glycine

had conventional voltage relations and were insensitive to 2-amino-5-phosphonovaleric acid (AP5) ($n=10/10$); they did not seem to contain a NMDA-receptor mediated component (Fig. 2h, k, l). In contrast, the 5 e.p.s.ps that were augmented by glycine had voltage relations that were typical of neuronal responses to NMDA^{1,2} (Fig. 2f) and were depressed by AP5 ($n=3/5$) (Fig. 2a–e). Although glycine produced a relatively small increase in the total e.p.s.p., it produced a substantial increase in the AP5-sensitive, NMDA receptor-mediated, component (Fig. 2g). Although it is possible that AP5 applied electrophoretically did not gain access to all NMDA receptors mediating the e.p.s.p., it is probable that a similar population of receptors was affected by both glycine and AP5, because we used submaximal doses of both drugs (assessed by their actions on NMDA responses) and there was no augmentation by glycine in the presence of AP5.

These results indicate that glycine can selectively enhance both neuronal responses to NMDA and NMDA-receptor-mediated

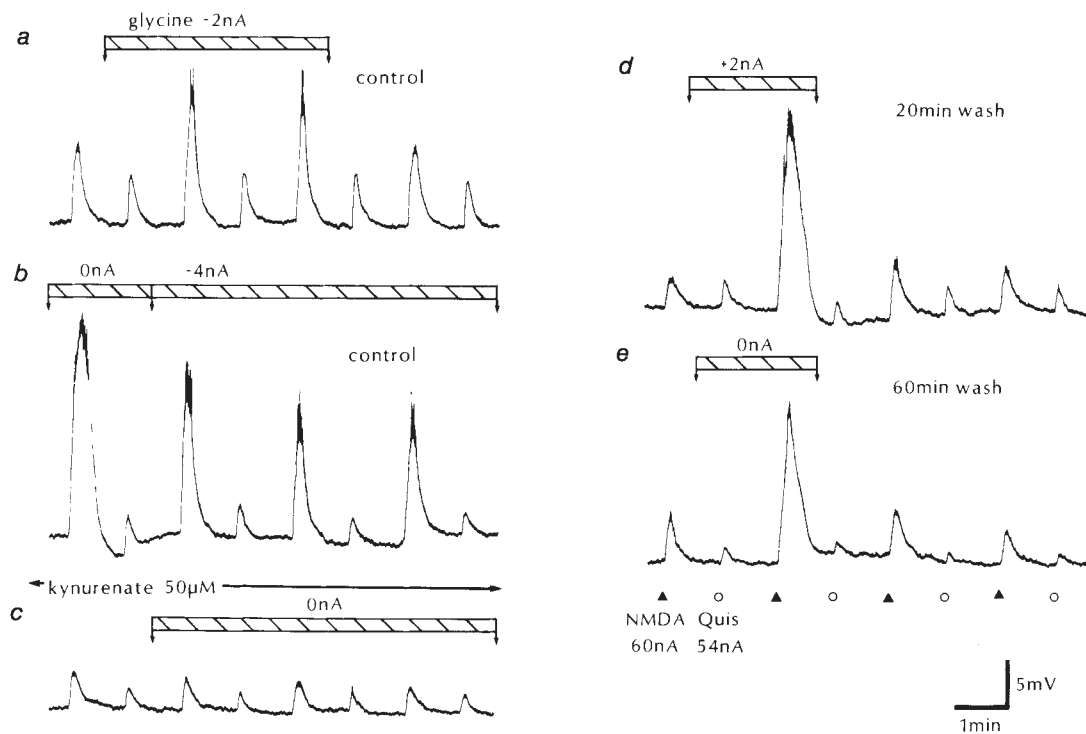


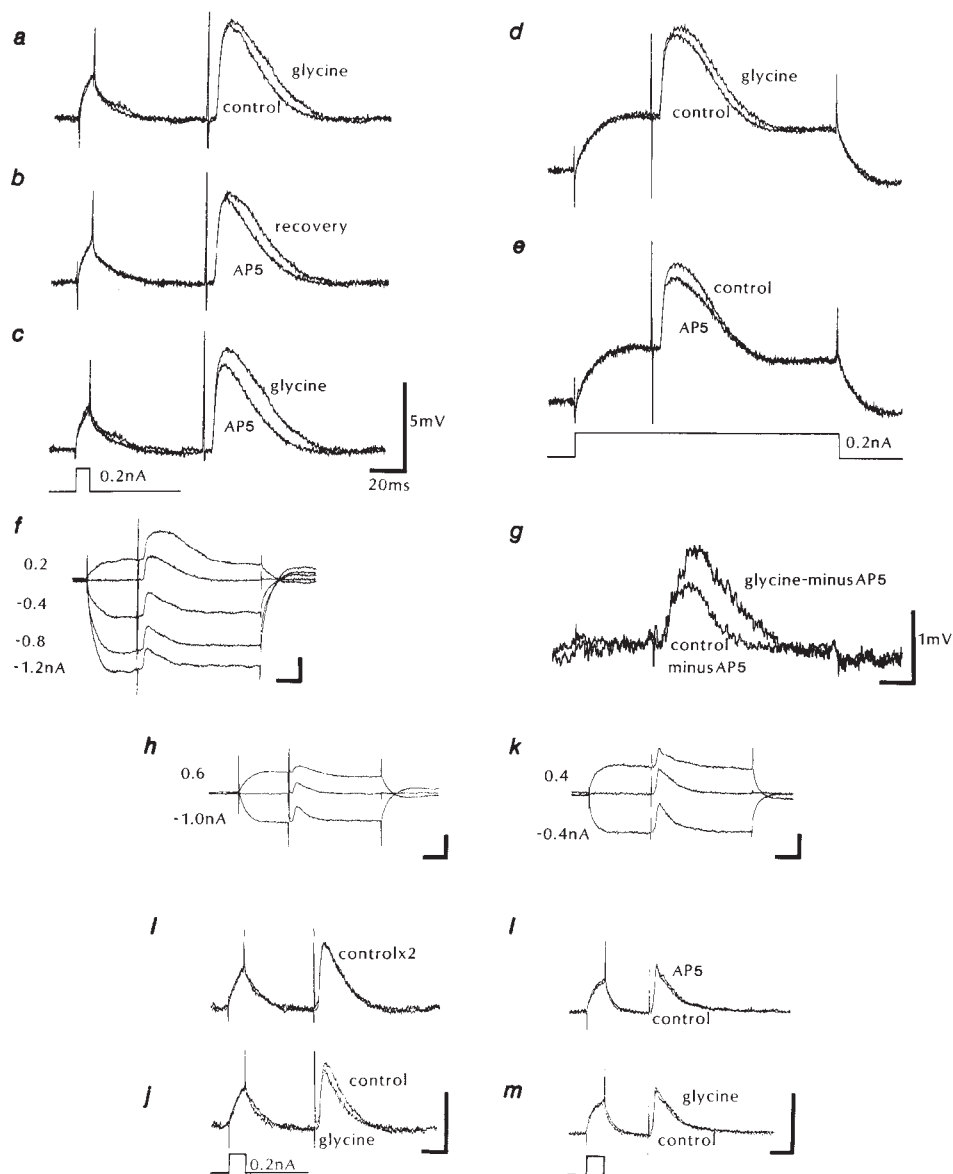
FIG. 1 Augmentation of responses to NMDA by glycine and blocking of the augmentation by kynurenate. Depolarizations of a cortical neuron in response to brief electrophoretic pulses of NMDA (closed triangles) and quisqualate (open circles), recorded intracellularly with conventional current-clamp techniques. The symbols in c and e apply to all records. Amino-acid currents were: Quisqualate, 54 nA, 2s; and NMDA, 60 nA, 2s throughout. In a, electrophoretic application of glycine (hatched bar) with a reduced retaining current (–2 nA) increased the response to NMDA, but not to quisqualate. In b, a larger application of glycine (0 nA retaining current) resulted in a larger augmentation. When the retaining current was increased to –4 nA, responses declined, but remained augmented for three further cycles until the retaining current was returned to its control value (–30 nA) when responses returned, on recovery, to pre-glycine amplitudes (data not shown). After six cycles of enhancement and recovery, kynurenate was applied in the bathing medium (50 μ M). In c, glycine applied 25 min after kynurenate addition, with no retaining current, did not augment NMDA responses. In d, after a 20-min washout, 0 nA glycine was still ineffective (data not shown) but a +2 nA ejection current produced enhancement. In e, after a 60-min washout, 0 nA glycine caused enhancement. All records were obtained at the resting potential –72 mV ($\pm$ 2 mV). Spikes were curtailed by the pen recorder response frequency. A gradual decline in responses to both amino acids occurred

during the 3-h recording period.

METHODS. Coronal slices of cingulate/sensorimotor cortex were obtained from adult female Wistar (12) and male Sprague–Dawley (6) rats, as described previously². Intracellular recordings were made from pyramidal neurons in layers II/III. Agonists and antagonists were applied electrophoretically from a 5-barrel pipette whose tip was within 50 μ m of the recording site. The barrels contained: NMDA, quisqualate (both 200 mM, pH 8), glycine (200 mM, pH 3 in 30 cells; 50 mM, pH 3 in 3 cells; 50 mM, pH 8 in 5 cells) and kynurenate (50 mM, pH 8) or D-AP5 (5 mM, pH 8) and NaCl (200 mM, balance barrel). Barrel resistances ranged from 70 to 120 M Ω in pipettes with which significant augmentation was achieved. Excitatory amino acids were applied in brief (1–2s) pulses (5–60 nA), repeated in an electronically controlled cycle with interpulse intervals of between 20 and 90 s. Glycine was applied continuously for 1 to 30 cycles either by reducing the retaining current (to 0–6 nA), or with low ejection currents (0 to 2 nA). Retaining currents for glycine of less than 10 nA resulted in a gradual loss of the augmentation of NMDA by glycine. Kynurenate ejection currents of between 2 and 20 nA rapidly and reversibly blocked the augmentation of NMDA responses by glycine. Larger currents and bath applications, of between 50 and 100 μ M reduced responses to both NMDA and quisqualate. In one experiment (three cells) strychnine (5–15 μ M) was bath-applied.

FIG. 2 E.p.s.ps recorded intracellularly in three cortical neurons (*a* to *g*, *h* to *j* and *k* to *m*). In *a*, the e.p.s.p. is augmented by glycine application (0 nA). After recovery (*b*), the e.p.s.p. is depressed by AP5 (10 nA). This dose of AP5 depressed responses to NMDA by more than 80%, but did not reduce quisqualate responses. The e.p.s.ps recorded in the presence of glycine and of AP5 are superimposed in *c*; for comparison. In *a*, *b*, *c*, small injected current pulses (0.2 nA), see lower record in *c*, precede stimuli. Note the lack of effect of either AP5 or glycine on the rising phase of the e.p.s.p., which may be due to activation of non-NMDA receptors. We repeated this procedure (*d*, *e*) but with e.p.s.ps superimposed on longer injected current pulses (0.2 nA, lower record in *e*). These responses were then digitally subtracted and amplified to reveal the AP5-sensitive component under control and glycine conditions (superimposed in *g*). *f*, The unconventional voltage relation of this e.p.s.p., which is typical of responses mediated at least in part by NMDA receptors^{1-3,12}. Stimuli were superimposed on injected current pulses (amplitude of pulses given to the left). *h*, *k*. The conventional voltage relations of two other e.p.s.ps. Three control averages of the e.p.s.p. shown in *h* are illustrated in *i* and in the control record *j*. This e.p.s.p. was depressed by glycine (10 nA ejection current) and there was a concurrent, if smaller, decrease in the voltage response to current injection. The e.p.s.p. shown in *k* was not significantly reduced by AP5 (*l*) nor augmented by glycine (2 nA ejection current; *m*). All records were obtained from resting potential: *a* to *g* -70 mV, *h* to *j* -72 mV, *k* to *m* -79 mV. All scale bars: 5 mV (vertical) and 20 ms (horizontal) except for *g* in which the vertical scale bar is 1 mV. *a*-*e*, *g*, *i*, *j*, *k*, *l*, *m* are averages each of 16 sweeps. *f*, *h* and *k* are averages each of eight sweeps.

METHODS. In six experiments, electrical stimulation ($<1 \mu\text{A}$, $<0.2 \text{ ms}$, 1 every 5 s) of the underlying white matter was used to evoke small e.p.s.ps. E.p.s.ps and voltage responses to small, brief, injected current pulses were averaged (eight to 16 sweeps). Any given response was considered acceptable when three successive averages were similar in shape and amplitude. Augmentation of e.p.s.ps by glycine, or their depression by AP5 were accepted only when responses recovered to pre-



drug levels and the effect could be repeated. For example, the e.p.s.p. illustrated in *a* to *g* was exposed to three glycine and two AP5 applications and recovered from each. Voltage relations for evoked e.p.s.ps were obtained by superimposing e.p.s.ps on longer duration injected current pulses.

ated e.p.s.ps in adult neocortical slices. As the pyramidal neurons of the cortex receive many conventional, non-NMDA-receptor-mediated inputs, the contribution made by NMDA receptors to their total excitatory input is very small. Even those e.p.s.ps that are mediated by NMDA receptors usually contain a non-NMDA-receptor-mediated component³ (see also ref. 12). It is therefore unlikely that changes in local glycine concentration will dramatically change the overall excitatory drive to these cells. But because NMDA receptors are involved in neuronal plasticity, the present demonstration that NMDA-receptor-mediated responses can be enhanced by exogenous glycine indicates that glycine has an important modulatory role. □

Received 5 January; accepted 8 February 1989.

1. Thomson, A. M., West, D. C. & Lodge, D. *Nature* **313**, 479-481 (1985).
2. Thomson, A. M. *J. Physiol.* **370**, 531-549 (1986).
3. Thomson, A. M., Girdlestone, D. & West, D. C. *B. J. Pharmac.* **96**, 406-408 (1989).
4. Rauschecker, J. P. & Hahn, S. *Nature* **326**, 183-185 (1987).
5. Artola, A. & Singer, W. *Nature* **330**, 649-652 (1987).
6. Morris, R. G. M., Anderson, E., Lynch, G. S. & Baudry, M. *Nature* **319**, 774-776 (1986).
7. Johnson, J. W. & Ascher, P. *Nature* **325**, 529-531 (1987).
8. Bonhaus, D. W. & McNamara, J. O. *Molec. Pharmac.* **34**, 250-255 (1988).
9. Fletcher, E. J. & Lodge, D. *Eur. J. Pharmac.* **151**, 161-162 (1988).
10. Kessler, M., Baudry, M., Terramani, T. & Lynch, G. *Soc. Neurosci. Abs.* **13**, 760 (1987).
11. Araki, T. *et al. Neurosci.* **25**, 613-624 (1988).
12. Forsythe, I. D. & Westbrook, G. L. *J. Physiol.* **396**, 515-533 (1988).

ACKNOWLEDGEMENTS This work was supported by the Wellcome Trust and the MRC.

Regulation of NMDA receptor desensitization in mouse hippocampal neurons by glycine

Mark L. Mayer, Ladislav Vyklicky Jr & John Clements

Unit of Neurophysiology & Biophysics, Laboratory of Developmental Neurobiology, NICHD, National Institute of Health, Bethesda, Maryland 20892, USA

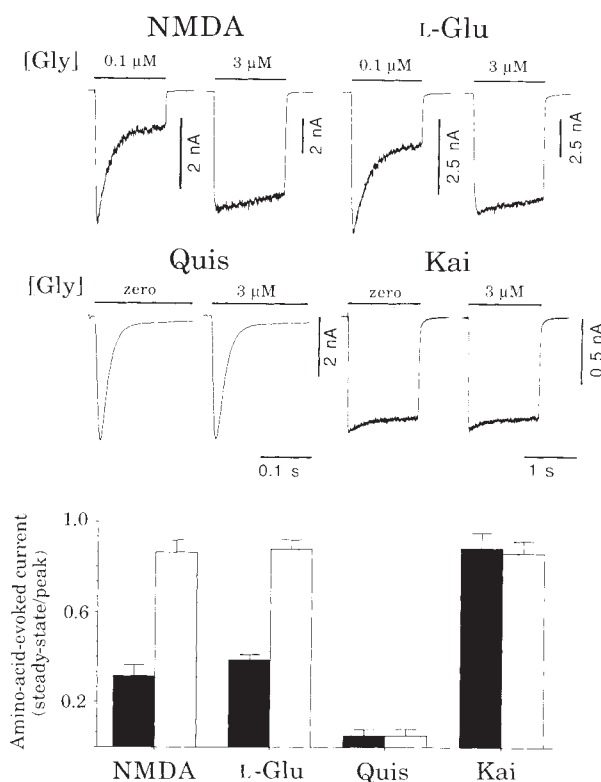
RESPONSES to the excitatory amino acid *N*-methyl-D-aspartate (NMDA) are markedly potentiated by nanomolar concentrations of glycine¹⁻³. This is due to the action of glycine at a novel strychnine-resistant binding site with an anatomical distribution identical to that for NMDA receptors^{4,5}, suggesting that the NMDA receptor channel complex contains at least two classes of amino-acid recognition site. Antagonists at the glycine-binding site associated with NMDA receptors act as potent non-competitive antagonists^{6,7}, but do not alter the mean open time or conductance, as estimated by fluctuation analysis⁸. The mechanisms by which glycine acts on NMDA receptors are unknown, but single-channel recording experiments show an increase in opening frequency with no change in mean open time or conductance⁷, suggesting that glycine could regulate transitions to states that are intermediate

between binding of NMDA receptor agonists and ion-channel gating. It has been suggested that glycine acts as a co-agonist at the NMDA receptor, and that responses to NMDA cannot be obtained in the complete absence of glycine⁹, but in these experiments the response to NMDA was measured at equilibrium, and it is unlikely that sufficient temporal resolution was achieved to detect rapid alterations in receptor gating. Using a fast perfusion system we find that glycine regulates desensitization at NMDA receptors; this has a major effect on the response to NMDA measured at equilibrium, as would occur with slower applications of agonist. Reduction of NMDA receptor desensitization by glycine provides an example of a novel mechanism for regulation of ion-channel activity.

We observed that fast applications of excitatory amino acids to hippocampal neurons in dissociated culture produced three patterns of response consistent with activation of kainate, NMDA and quisqualate receptors¹⁰: responses to kainate showed little or no desensitization; responses to NMDA desensitized relatively slowly with a time constant of around 250 ms; responses to quisqualate desensitized very rapidly, with a time constant of around 10–20 ms or less. In subsequent experiments (Fig. 1) we found that desensitization at NMDA receptors, but not at kainate or quisqualate receptors, was sensitive to changes in the concentration of glycine, such that the desensitization of responses to NMDA was greatly reduced with 3 or 10 μ M glycine. Similar effects were observed with activation of NMDA receptors by both NMDA and L-glutamate, whereas D-serine and D-alanine (3 μ M) were able to substitute for glycine in preventing desensitization.

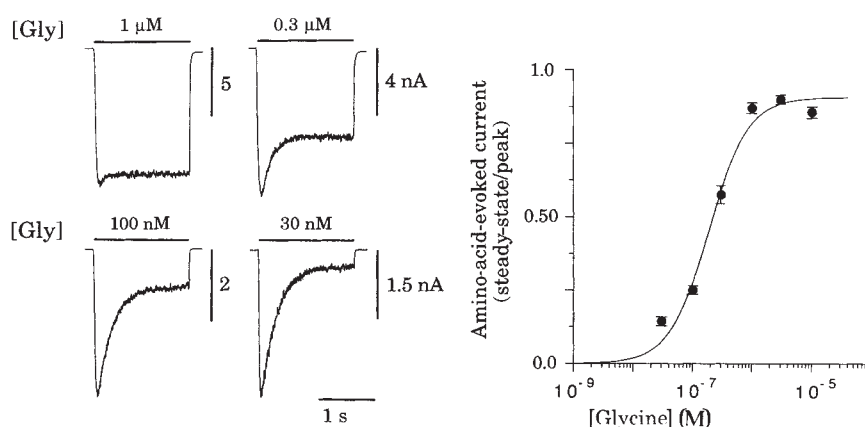
FIG. 1 Glycine regulates desensitization at NMDA receptors, but not at kainate (Kai) or quisqualate (Quis) receptors. Responses were evoked by fast applications of NMDA (100 μ M), L-glutamate (10 μ M), quisqualate (40 μ M) and kainate (50 μ M) with 0.1 μ M and 3 μ M glycine (NMDA and L-glutamate) or zero and 3 μ M glycine (quisqualate and kainate) using whole-cell recording under discontinuous voltage clamp. The responses for NMDA and L-glutamate recorded with 0.1 μ M glycine are plotted at twice the gain of the responses recorded with 3 μ M glycine, to emphasize the effect of glycine on desensitization. The time base for responses evoked by quisqualate is 10 times faster than for the other agonists. The onset of desensitization of the responses to NMDA and L-glutamate recorded with 0.1 μ M glycine could be fitted to exponential functions of time constant 220 and 305 ms respectively. The onset of desensitization of the response to quisqualate did not change with zero or 3 μ M glycine, and could be fitted to an exponential function of time constant 17.6 ms. The responses to NMDA and L-glutamate were recorded in the same neuron, and emphasize the difference in potency of L-glutamate and NMDA, which was confirmed in separate experiments (not shown) in which complete concentration response curves were obtained for each agonist; at these doses the responses to NMDA and L-glutamate are near maximum. The residual desensitization of responses to NMDA recorded with 3 μ M glycine was fairly variable and essentially absent in some experiments; as this process was accelerated on raising the extracellular calcium concentration, or on substituting EGTA for BAPTA as an intracellular calcium buffer, we suggest that it represents a separate process, triggered by a rise in intracellular calcium ion concentration as a result of calcium influx through NMDA receptor channels. The bar graph summarizes results from experiments on nine neurons, and is plotted as steady state (measured at 1.5 s)/peak current; error bars show 1 standard deviation. Responses recorded with 0.1 μ M glycine are represented by filled columns; responses with 3 μ M glycine by open columns.

METHODS. Experiments used primary cultures of hippocampal neurons from 16–17-day-old mouse embryos plated on a primary culture of glial cells also from mouse hippocampus¹⁵. Neurons were kept in culture for up to three weeks, but most experiments were performed 4–14 days after plating, as development of the dendritic field made rapid solution changes more difficult to attain in older cells. In experiments in which cells were bathed in 50 μ M kainic acid and the extracellular sodium concentration stepped from 5 to 165 mM, an inward relaxation developed with a time constant of 7–10 ms indicating that there was rapid perfusion under our experimental conditions. The perfusion apparatus was similar to that used by Johnson and Ascher^{11,12} and is described in more detail elsewhere¹⁰. The extracellular solution contained (in mM): 165 NaCl, 2.5 KCl, 2 CaCl₂, 1 MgCl₂, 10 glucose, 10 HEPES, pH 7.3, with 400 nM TTX (tetrodotoxin) and 5 μ M bicuculline methiodide,



and glycine added as indicated. In all experiments, neurons were pre-equilibrated with the required concentration of glycine for 10–20 s before NMDA was applied together with glycine. NMDA and L-glutamate were applied using a modified extracellular solution containing 0.2 mM CaCl₂ to reduce calcium-mediated inactivation of NMDA receptor activity^{16,17} and no added Mg^{17,18} to prevent channel block. The intracellular solution contained (in mM): 125 CsMeSO₃, 15 CsCl, 5 CsBAPTA (1,2-bis(o-aminophenoxy) ethane-N,N,N',N'-tetraacetic acid), 0.5 CaCl₂, 5 MgCl₂ and 2 ATP. Experiments were performed at room temperature (25–27 °C).

FIG. 2 Concentration-dependent regulation by glycine of NMDA receptor desensitization. Traces show the responses to rapid applications of 100 mM NMDA at the glycine concentrations indicated and are plotted so that the peak amplitude of the traces is similar. Desensitization is pronounced at low concentrations of glycine, but is essentially absent at high concentrations of glycine. Note the similar time course of the onset of desensitization at different concentrations of glycine. Responses to 30 nM and 100 nM glycine were recorded from the same neuron, responses to 0.3 μ M and 1 μ M glycine were recorded from a second neuron. The dose-response curve summarizes results from 13 neurons each of which was tested with four concentrations of glycine. The y axis is a measure of desensitization and shows values calculated as steady-state current (measured at 1.5 s) divided by the peak-current response to NMDA. The data points are fitted by nonlinear regression to the Hill equation (maximum=0.9; K_d =185 nM; n =1.3; error bars show 1 s.d. Further analysis of these experiments showed potentiation by glycine of both the peak and steady-state response to NMDA, although because of the effect of glycine on desensitization the potentiation of the steady-state response to NMDA was much larger: on average, the peak



current increased by 3.53 on raising the glycine concentration from 30 nM to 10 μ M glycine, whereas the steady state current increased by 26.6. One interpretation of this result would be an agonist-evoked transition of the NMDA receptor to a state of lower affinity for glycine, such that in the presence of NMDA or L-glutamate, higher concentrations of glycine are required to maintain ion-channel activity.

Johnson and Ascher have described quantitative experiments on NMDA receptor modulation by glycine in mammalian central nervous system neurons, which suggest that glycine binds to a single site with an affinity constant of about 300 nM¹¹; similar estimates have been obtained using binding of [³H]glycine to NMDA receptors in membrane preparations⁸ or the application of NMDA receptor agonists to mRNA-injected oocytes under voltage clamp^{2,3,9}. In our experiments, the potentiation by glycine of NMDA receptor responses measured at steady state (at the end of 1.5-s applications of NMDA) could be described by similar parameters: a 24-fold maximum potentiation of responses to NMDA with 10 μ M glycine compared with responses measured with 3 nM glycine, with a dissociation constant of 360 nM and n =1.3, estimated using nonlinear regression to fit the Hill equation to data recorded from 13 neurons (data not shown).

To determine whether glycine regulation of NMDA receptor desensitization occurs over a similar concentration range, we examined responses to rapid applications of NMDA in neurons pre-equilibrated for 10–20 s with 3 nM–10 μ M glycine (Fig. 2). We used this procedure to allow pre-equilibration of glycine binding, as at sub-micromolar concentrations, diffusion-limited binding of glycine would be expected to markedly slow the interaction of glycine with NMDA receptors if a rapid change in the concentration of both glycine and NMDA was made simultaneously. Our results indicate that the glycine-binding site regulating desensitization of NMDA receptors is titrated over a similar concentration range to that regulating potentiation of responses to NMDA measured at equilibrium, as the dose-response analysis of the degree of desensitization of NMDA receptor responses could be described by a binding model with a dissociation constant of 185 nM, n =1.3. The requirement that applications of agonist are fast to allow recording of desensitization at NMDA receptors is revealed by inspection of responses recorded at low glycine concentrations ($\leq$ 100 nM) as, with 100 μ M NMDA, equilibrium desensitization (measured as the ratio of steady-state/peak current) is more than 75%, and its onset is rapid, being 90% complete within 500 ms.

Our results are in reasonable agreement with the original experiments of Johnson and Ascher¹ as, at the single-channel level, a reduction of desensitization by glycine would probably be recorded as an increase in frequency of ion-channel opening evoked by NMDA receptor agonists if the mean lifetime of desensitized states is long relative to the open and closed time distributions measured during normal gating behaviour of the NMDA receptor, as for nicotinic receptors¹². The analysis of

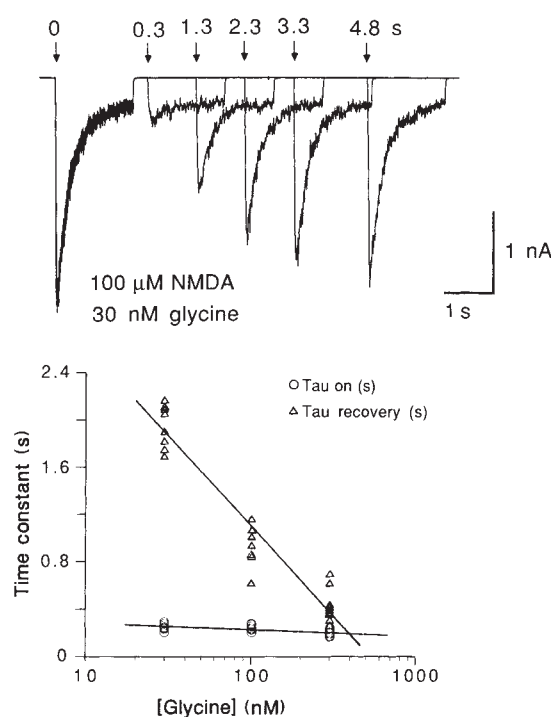


FIG. 3 Glycine increases the rate of recovery from desensitization at NMDA receptors. The traces show a family of currents evoked by paired pulse applications of 100 μ M NMDA with 30 nM glycine, recorded while the interval between the first and second pulse of NMDA was increased as indicated, to measure the speed of recovery from desensitization. In all cells studied with 30 nM glycine, complete recovery from desensitization required more than five seconds. When the peak amplitude of the response to the second application of NMDA, normalized with respect to the peak amplitude of the first response to NMDA, was plotted as a function of time, recovery from desensitization could be described by a single exponential function (tau recovery) which became faster as the concentration of glycine was increased; in the example shown, the time constant of recovery from desensitization was 3.1 s. The onset of desensitization could also be fitted by a single exponential function (tau on) but the time constant was much less sensitive to the concentration of glycine; in the traces shown, desensitization developed with a time constant of 277 ms. A graph showing the results of exponential fits to the onset and recovery from NMDA receptor desensitization in 25 experiments similar to that illustrated above is shown for 30, 100 and 300 nM glycine, and clearly shows the effect of glycine on the rate of recovery from desensitization, with minimal change in the time of onset of desensitization.

whole-cell currents provides one way to examine this possibility; we used paired pulse applications of NMDA at several concentrations of glycine to measure the rates of onset and recovery from desensitization as a function of glycine concentration (Fig. 3). Over the range 30 to 300 nM glycine, there is little change in the rate of onset of desensitization, but a dramatic speeding up of the rate of recovery from desensitization.

Although previous experiments have used fast applications of NMDA to study modulation by glycine, changes in desensitization were not reported²; this probably reflects application of agonist from a point source to a large area of membrane (*Xenopus laevis* oocytes used for voltage-clamp experiments are typically ≥ 1 mm in diameter), making it very difficult to achieve a rapid and uniform application of agonist, thus masking desensitization at NMDA receptors. Second, we have found that calcium influx through NMDA receptors triggers secondary mechanisms, which leads to glycine-resistant desensitization; to avoid this, we used a low extracellular concentration of calcium in all our experiments.

Further analysis of the mechanism of action of glycine on NMDA receptor channels will ultimately require a much deeper understanding of the receptor-gating process than has been achieved for any agonist-activated channel. Nonetheless, our results indicate that a major component of the action of glycine on NMDA receptors involves speeding up the rate constant of recovery from desensitization, such that NMDA receptor channels can open more frequently at equilibrium because the rate of escape from a long-lived desensitized state increases as the concentration of glycine is raised. To our knowledge, these are the first results describing regulation of ion-channel activity by an allosteric mechanism involving reduction of desensitization. Indeed, an initial inspection of the common models for desensitization at the much better characterized nicotinic acetylcholine receptor³ suggests that processes that increase the degree of occupation of the agonist recognition site or the frequency of ion-channel activation would be expected to increase desensitization, as binding of agonist, receptor activation and desensitization are all tightly coupled processes. Experiments on GABA_A (gamma-aminobutyric acid) receptor channels have shown that the benzodiazepine modulator chlordiazepoxide, which enhances responses to GABA, also increases the degree of desensitization at any given dose of GABA¹⁴. In contrast, modulation of desensitization at nicotinic acetylcholine receptors by substance P occurs without any change in the amplitude of the initial response to agonist²⁰ and differs from the effect of glycine described here in that it occurs by acceleration of the onset of desensitization, with no change in the rate of recovery from desensitization. As a result, substance P markedly inhibits responses to acetylcholine measured at equilibrium. Together, the effects of glycine and substance P provide dramatic examples of how modulation of desensitization can regulate the response to neurotransmitters.

Whether desensitization at NMDA receptors occurs under physiological conditions will be determined by the time course of the excitatory synaptic transmitter concentration transient at synaptic boutons. The decay time constant of the NMDA receptor component of synaptic current is unusually long and can last for hundreds of milliseconds²¹, which is within the period over which desensitization occurs in biophysical experiments of the type described here. The long time course for the synaptic current probably reflects the continued presence of transmitter, as we have found rapid deactivation of NMDA receptor current (time constant ≤ 10 ms) on removal of agonist in concentration jump experiments. This suggests that the regulation by glycine of NMDA receptor desensitization *in vivo* could have important functional consequences. □

3. Kushner, L., Lerma, J., Zukin, S. R. & Bennett, M. V. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3250–3254 (1988).
4. Bristow, D. R., Bowery, N. G. & Woodruff, G. N. *Eur. J. Pharmac.* **126**, 303–307 (1986).
5. Monaghan, D. T. & Cotman, C. W. *J. Neurosci.* **5**, 2909–2919 (1986).
6. Evans, R. H., Evans, S. J., Pook, S. C. & Sunter, D. C. *Br. J. Pharmac.* **91**, 531–537 (1987).
7. Kemp, J. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6547–6550 (1987).
8. Mayer, M. L., Westbrook, G. L. & Vyklicky, L. *J. Neurophysiol.* **60**, 645–663 (1988).
9. Kleckner, N. W. & Dingledine, R. *Science* **241**, 835–837 (1988).
10. Mayer, M. L. & Vyklicky, L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1411–1415 (1989).
11. Johnson, J. W. & Ascher, P. *Soc. Neurosci. Abstr.* **13**, 383 (1987).
12. Sakmann, B., Patlak, J. & Neher, E. *Nature* **286**, 71–73 (1980).
13. Feltz, A. & Trautmann, A. *J. Physiol.* **322**, 257–272 (1982).
14. Mierlak, D. & Farb, D. H. *J. Neurosci.* **8**, 814–820 (1988).
15. Mayer, M. L. & Westbrook, G. L. *J. Physiol.* **394**, 501–527 (1987).
16. Mayer, M. L. & Westbrook, G. L. *J. Physiol.* **361**, 65–90 (1985).
17. Mayer, M. L., MacDermott, A. B., Westbrook, G. L., Smith, S. J. & Barker, J. L. *J. Neurosci.* **7**, 3230–3244 (1987).
18. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462–465 (1984).
19. Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261–263 (1984).
20. Clapham, D. & Neher, E. *J. Physiol.* **347**, 255–277 (1984).
21. Forsythe, I. D. & Westbrook, G. L. *J. Physiol.* **396**, 515–533 (1988).

ACKNOWLEDGEMENTS. We thank Sandy Fitzgerald for preparing cultures, and Drs B. Smith and S. Hsiao for help with constructing the perfusion system. L.V. was supported by a Fogarty fellowship.

Cytosolic calcium regulates ion channels in the plasma membrane of *Vicia faba* guard cells

Julian I. Schroeder & Susumu Hagiwara

Department of Physiology, Jerry Lewis Neuromuscular Research Center, UCLA School of Medicine, Los Angeles, California 90024, USA

THE molecular mechanisms by which Ca^{2+} controls a variety of ion transport-associated cellular functions in higher plant cells¹, including movements of stomatal pores, remain unknown. Stomatal pores regulate the gas exchange in leaves. Openings of stomatal pores are mediated by an increase in the intracellular potassium and anion content of guard cells^{2,3}. Voltage-dependent K^{+} channels^{4–7} and hyperpolarizing H^{+} pumps^{8,9} have been identified as mechanisms controlling stomatal movements. But depolarizing mechanisms required for K^{+} efflux during stomatal closing remain unknown. Indirect evidence suggests that Ca^{2+} triggers stomatal closing and inhibits stomatal opening^{10–12}. Using patch-clamp techniques¹³, we show here that elevated (micromolar) concentrations of cytosolic Ca^{2+} in guard cells block inward rectifying K^{+} channels. Furthermore, elevation of cytosolic Ca^{2+} leads to the activation of a voltage-dependent depolarizing conductance with a permeability to anions. The Ca^{2+} -induced modulation of ion channels reported here could provide a molecular basis for Ca^{2+} -dependent regulation of stomatal movements in leaves.

Figure 1a illustrates typical voltage-dependent inward and outward conducting K^{+} channel currents^{5,6} recorded across the plasma membrane (plasmalemma) of a guard cell using the whole-cell patch clamp technique¹³. When the cytoplasmic Ca^{2+} concentration ($[\text{Ca}^{2+}]_{\text{cyt}}$) was buffered to low values of 0.01 μM and 0.1 μM , the inward conducting K^{+} current ($I_{\text{K}^{+},\text{in}}$) was activated at potentials more negative than -87 mV (± 12 mV, $\pm$ s.e.m. $n = 12$) (Fig. 1a and c).

Raising the cytosolic Ca^{2+} concentration to 1.5 μM resulted in several dramatic changes in the guard cell membrane currents (Fig. 1b). The inward rectifying K^{+} current was inhibited in the normal activation range when $[\text{Ca}^{2+}]_{\text{cyt}}$ was elevated (Fig. 1b, c). With 1.5 μM $[\text{Ca}^{2+}]_{\text{cyt}}$, time-dependent inward rectifying currents were activated at potentials more negative than -192 mV (± 13 mV, $\pm$ s.e.m., $n = 19$) (Fig. 1b). When $[\text{Ca}^{2+}]_{\text{cyt}}$ was buffered to 9 μM , inward current activation occurred at potentials more negative than -196 mV (± 13 mV, $n = 11$). Tail currents⁵ of these hyperpolarization-induced currents at elevated $[\text{Ca}^{2+}]_{\text{cyt}}$ reversed at the K^{+} equilibrium potential, indicating that they

Received 9 December 1988; accepted 15 February 1989.

1. Johnson, J. W. & Ascher, P. *Nature* **325**, 529–531 (1987).
2. Verdoorn, T. A., Kleckner, N. W. & Dingledine, R. *Science* **238**, 1114–1116 (1987).

were carried by K^+ ions (data not shown). Thus, elevation of cytosolic Ca^{2+} seems to shift the activation potential of $I_{K^+,in}$ to large negative potentials.

To test this suggestion, single-channel currents were recorded in cell-free outside-out membrane patches¹³ (Fig. 2). With low free Ca^{2+} concentrations ($0.01 \mu M$) on the cytoplasmic side of the membrane, increasingly negative potentials enhanced the probability of channel opening (Fig. 2a). But when $[Ca^{2+}]_{cyt}$ was elevated, only few discrete single-channel current steps were observed at large negative potentials (Fig. 2b).

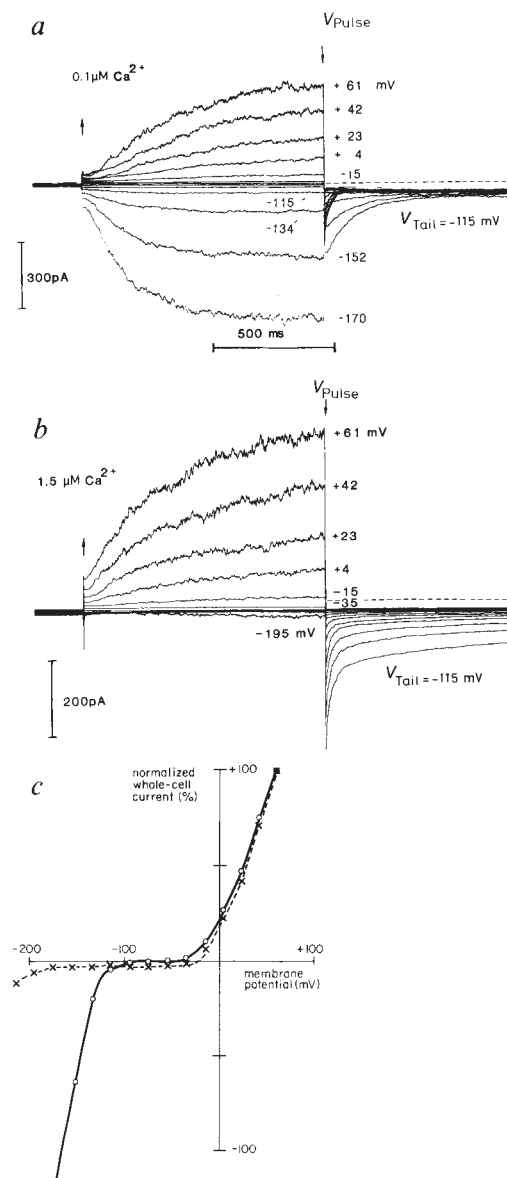
Single-channel conductances recorded at low cytoplasmic Ca^{2+} ranged from 5 to 8 pS ($10^{-12} \Omega^{-1}$, $n = 3$). Channel conductances recorded at elevated $[Ca^{2+}]_{cyt}$ were 4–7 pS ($n = 3$). Extrapolation of single-channel current amplitudes with low and elevated $[Ca^{2+}]_{cyt}$ resulted in reversal potentials close to the K^+ -equilibrium potential ($E_{K^+} = -54$ mV). These results indicate that the Ca^{2+} -dependent reduction of $I_{K^+,in}$ was mediated by a shift in the activation potential of $I_{K^+,in}$ channels to large negative potentials. Moreover, experiments with excised membrane patches demonstrate that cytoplasmic Ca^{2+} modulates $I_{K^+,in}$ channels by directly effecting plasma membrane-bound processes, rather than by interacting with other intracellular factors.

FIG. 1 Effect of cytosolic Ca^{2+} on guard cell plasma membrane currents in whole-cell recordings. **a**, Inward conducting K^+ currents^{4–6} (downward deflections) and outward conducting K^+ currents^{5–7} (upward deflections) recorded with $0.1 \mu M$ free cytosolic Ca^{2+} . Consecutive one-second voltage steps were applied from a holding potential of -75 mV to pulse potentials (V_{Pulse} , between the arrows) ranging from $+61$ mV to -170 mV, as indicated to the right of currents. After V_{Pulse} , the membrane was stepped to -115 mV (V_{Tail}). (The group of rapidly deactivating currents at $V_{Tail} = -115$ mV corresponded to deactivation of $I_{K^+,out}$, whereas the group of slowly deactivating currents corresponded to deactivation of $I_{K^+,in}$.) Whole-cell capacitance was 5.35 pF. Membrane potential was -52 mV (zero current). **b**, Whole-cell currents measured as described in **a**, with $1.5 \mu M$ free Ca^{2+} in the cytoplasm. The membrane potential was -20 mV (zero current). The zero-current level is indicated by the broken line on the right in **(a)** and **(b)**. Whole-cell capacitance was 5.57 pF. **c**, Normalized whole-cell currents from cells shown in **(a)** and **(b)** plotted as a function of V_{Pulse} (open circles: $[Ca^{2+}]_{cyt} = 0.1 \mu M$; crosses: $[Ca^{2+}]_{cyt} = 1.5 \mu M$).

METHODS. Guard cell protoplasts were isolated from the lower leaf epidermis of *Vicia faba* as previously described⁶ and were bathed during recordings in external solution containing: 10 mM potassium glutamate, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM KOH, 10 mM MES 2-[*N*-morpholino]ethanesulphonic acid, pH 5.5, osmolality adjusted to 480 mmol kg^{-1} with D-mannitol. The internal solution, which equilibrated with the cytoplasm, contained: 100 mM potassium glutamate, 2 mM $MgCl_2$, 4 mM KOH, 10 mM HEPES, 4 mM MgATP, pH 7.2, and D-mannitol, osmolality 530 mmol kg^{-1} . Concentrations of free Ca^{2+} were buffered in the internal solution by addition of 0.45 mM $CaCl_2$ and 1.1 mM K_2 -EGTA in **(a)** and by addition of 1 mM $CaCl_2$ and 2 mM 5,5'-dibromoBAPTA, tetrapotassium (K_4 -D-Br-BAPTA) (Molecular Probes, Oregon) in **(b)**. Ca^{2+} -EGTA buffers (dissociation constant $K_D = 0.147 \mu M$) were prepared to compensate for impurities of EGTA³⁰. Ca^{2+} buffers using K_4 -D-Br-BAPTA ($K_D \sim 1.5 \mu M$)³⁰ were corrected for impurities and effects of ionic strength by calibration of internal solutions using Ca^{2+} -selective electrodes. Voltage-clamp and zero current-clamp recordings were performed with an EPC-7 patch-clamp amplifier (List Electronic) and analysed⁶ on a PDP 11/23 computer. The membrane was voltage-clamped to -75 mV for 12 s between consecutive voltage pulses. Liquid junction and series resistance potentials were accounted for⁶. All whole-cell recordings had effective access resistances⁵ of ≤ 10 M Ω to ensure efficient control of $[Ca^{2+}]_{cyt}$. The temperature was $22^\circ C$.

Elevation of $[Ca^{2+}]_{cyt}$ produced further changes in the ionic currents of guard cells. Depolarizing voltage steps with $1.5 \mu M$ $[Ca^{2+}]_{cyt}$ enhanced voltage-dependent instantaneous currents (Fig. 1b) in addition to the time-dependent K^+ currents observed at low $[Ca^{2+}]_{cyt}$, and repolarization of the membrane to -115 mV (V_{Tail}) evoked slow relaxation currents in addition to the rapidly deactivating $I_{K^+,out}$ tail currents (compare $V_{Tail} = -115$ mV in Fig. 1a, b). In current-clamp recordings of whole cells dialysed with low $[Ca^{2+}]_{cyt}$ ($\leq 0.1 \mu M$), the membrane potentials (V_m) were never more positive than the activation potential of $I_{K^+,out}$ ($V_m = -39 \pm 15$ mV, $\pm$ s.e.m., $n = 11$). Interestingly, elevation of cytosolic Ca^{2+} levels ($1.5 \mu M$ and $9 \mu M$) often resulted in membrane potentials that were more positive than the activation potential of whole-cell currents (Fig. 1b, c).

To examine the ion selectivity of the $[Ca^{2+}]_{cyt}$ -dependent depolarizing currents, solutions were prepared that would shift anion equilibrium potentials to more positive values and cation equilibrium potentials to more negative values (see Fig. 3 legend). Furthermore, to abolish K^+ channel currents, we replaced K^+ in the cytoplasmic and bath solutions by minutely permeant Cs^+ ions⁶. With $0.01 \mu M$ $[Ca^{2+}]_{cyt}$ in whole cells, only small currents were recorded (Fig. 3a), whereas Ca^{2+} -activated currents appeared with $1.5 \mu M$ $[Ca^{2+}]_{cyt}$ (Fig. 3b).



Using these Cs^+ solutions, the membrane potentials (V_m) of the cells with elevated $[\text{Ca}^{2+}]_{\text{cyt}}$ ($1.5 \mu\text{M}$) depolarized to $+15 \text{ mV}$ ($\pm 12 \text{ mV}$, $\pm \text{s.e.m.}$, $n = 7$). When Cs^+ ions were replaced by K^+ ions in the solutions used in Fig. 3, membrane potentials also depolarized with $1.5 \mu\text{M}$ $[\text{Ca}^{2+}]_{\text{cyt}}$ ($V_m = +9 \pm 8 \text{ mV}$, $\pm \text{s.e.m.}$, $n = 9$). But membrane potentials were substantially more negative when cytosolic Ca^{2+} was buffered to low concentrations of 0.01 and $0.1 \mu\text{M}$ ($V_m = -40.5 \pm 15 \text{ mV}$, $\pm \text{s.e.m.}$, $n = 15$).

Although $[\text{Ca}^{2+}]_{\text{cyt}}$ may also modulate outward K^+ currents, the Ca^{2+} -activated depolarizing currents can be attributed only to Ca^{2+} or anion permeation, as the equilibrium potentials of all other ions present were 0 mV , or more negative (Fig. 3 legend). Lowering external Ca^{2+} from 1 mM to 0.1 mM did not hyperpolarize the membrane or reduce currents, as seen in Fig. 3b. These results indicate that the Ca^{2+} -activated depolarizing currents were not carried by Ca^{2+} ions and suggest that membrane depolarizations were induced by a Ca^{2+} -activated anion permeability. This suggestion was supported by the finding that the membrane depolarized to between $+30 \text{ mV}$ and $+50 \text{ mV}$ when the cytoplasm was loaded with solutions containing 100 mM CsCl and cells were bathed in 10 mM CsCl ($n = 4$).

The Ca^{2+} -activated currents and the Ca^{2+} -induced depolarizations usually diminished within 2–10 min of establishing whole-cell recordings ('wash-out'). The 'wash-out' of the Ca^{2+} -dependent depolarizing currents indicated that other factors may be important for the activation of this conductance.

Our results directly demonstrate that variation of cytosolic Ca^{2+} in the range reported to be physiological (0.1 to $10 \mu\text{M}$)^{14–19}, strongly modulates ion channels in the plasma membrane of a higher plant cell. Extracellular manipulations of Ca^{2+} in stomata during physiological stimulation by light and abscisic acid^{10–12} have recently indicated that Ca^{2+} may exert two separate actions on stomatal movements: first, Ca^{2+} is required to initiate stomatal closing^{10–12} which relies on reduction of the K^+ and anion content of guard cells^{2,20}; and second, Ca^{2+} inhibits stomatal opening^{10–12} and the required K^+ influx¹¹.

Previous investigations have provided biophysical and pharmacological evidence that $I_{\text{K}^+, \text{in}}$ channels represent a major pathway for K^+ influx during stomatal opening^{4–6}. Thus, the inhibition of $I_{\text{K}^+, \text{in}}$ channels at elevated $[\text{Ca}^{2+}]_{\text{cyt}}$ reported here may provide a mechanism for the observed prevention of stomatal opening by Ca^{2+} . In animal cells, elevation of $[\text{Ca}^{2+}]_{\text{cyt}}$ can activate outward conducting K^+ channels^{21,22}. Interestingly, both the $[\text{Ca}^{2+}]_{\text{cyt}}$ -dependent inhibition of $I_{\text{K}^+, \text{in}}$ found here, and Ca^{2+} stimulation of K^+ channels in animal cells are mediated by hyperpolarizing shifts in the activation potential at increased $[\text{Ca}^{2+}]_{\text{cyt}}$ (Figs 1 and 2; ref. 22).

In addition to reducing $I_{\text{K}^+, \text{in}}$, elevation of $[\text{Ca}^{2+}]_{\text{cyt}}$ led to activation of a voltage-dependent depolarizing conductance

FIG. 2 Cytosolic Ca^{2+} modulates single $I_{\text{K}^+, \text{in}}$ channels in outside-out plasma membrane patches. The membrane potential of outside-out patches was continuously polarized to negative values at which the time-dependent activation of single $I_{\text{K}^+, \text{in}}$ channels has previously been described². a, Recordings of single K^+ channel currents at $0.01 \mu\text{M}$ $[\text{Ca}^{2+}]_{\text{cyt}}$. Increasingly negative holding potentials, as indicated to the left of the current traces, led to increased opening of $I_{\text{K}^+, \text{in}}$ channels. Small patch pipette tips were used to reduce the number of channels in the patch (tip resistance $10 \text{ M}\Omega$ in 150 mM KCl solutions⁴). b, Current recordings in an outside-out membrane patch with $9 \mu\text{M}$ free Ca^{2+} on the cytoplasmic membrane side. Cytosolic Ca^{2+} was buffered to $9 \mu\text{M}$ by addition of 0.66 mM CaCl_2 and 0.77 mM $\text{K}_4\text{-D-Br-BAPTA}$. Solutions are described in the legend to Fig. 1.

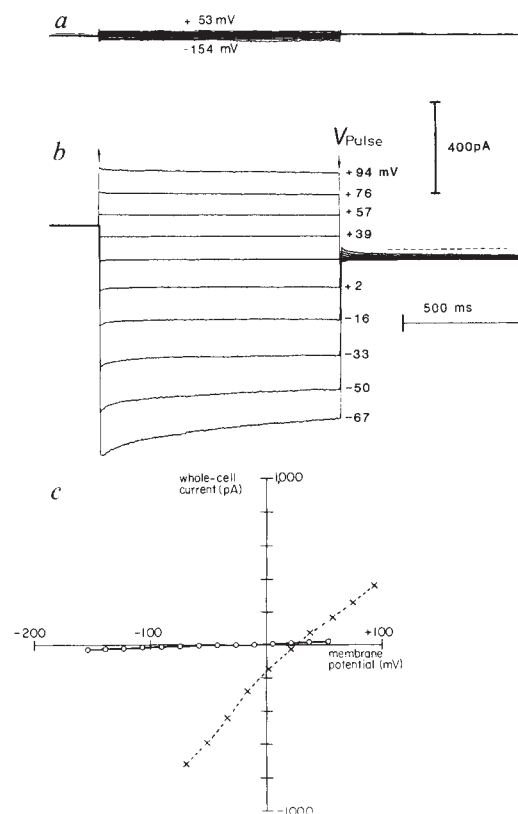
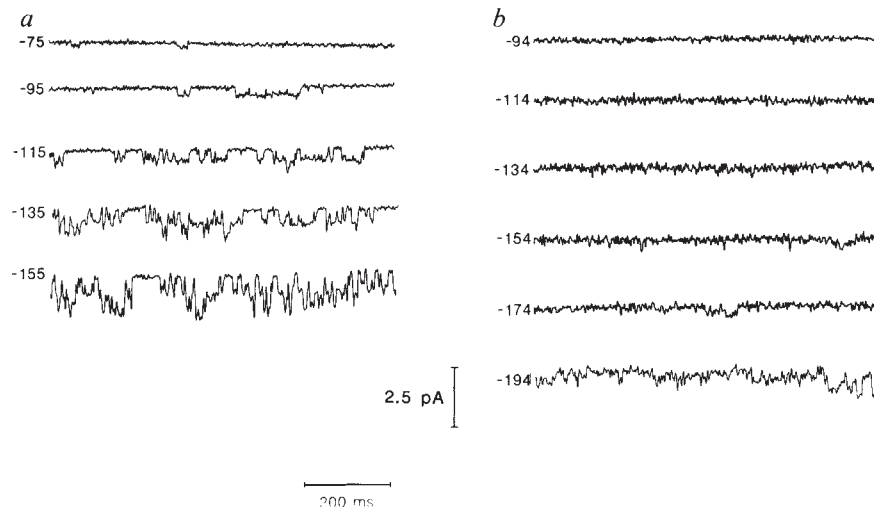


FIG. 3 Effect of cytosolic Ca^{2+} on whole-cell currents recorded with Cs^+ solutions. a, Whole-cell currents in response to voltage steps in the range from $+53 \text{ mV}$ to -154 mV with the cytoplasm exposed to $0.01 \mu\text{M}$ Ca^{2+} . b, Whole-cell currents in response to voltage steps from $+48 \text{ mV}$ with $1.5 \mu\text{M}$ free Ca^{2+} in the pipette solution. The zero-current level is indicated by the broken line on the right. In 84% of the guard cell protoplast preparations ($n = 25$) Ca^{2+} -activated currents as shown in Figs 1b and 3b were detected ('wash-out'). c, Current-voltage relationships of cells shown in (a) and (b) (open circles: $[\text{Ca}^{2+}]_{\text{cyt}} = 0.01 \mu\text{M}$; crosses: $[\text{Ca}^{2+}]_{\text{cyt}} = 1.5 \mu\text{M}$). The internal Cs^+ solution contained: 80 mM Cs-glutamate , 17.6 mM CsCl , 2 mM MgCl_2 , 4 mM MgATP , $\text{pH } 7.2$ and 1 mM total $\text{Cs}_2\text{-EGTA}$ in (a) and 1 mM CaCl_2 , 2 mM $\text{K}_4\text{-D-Br-BAPTA}$ in (b). The external Cs^+ solution contained: 8 mM Cs-glutamate , 0.5 mM CsCl , 1 mM CaCl_2 , 10 mM HEPES , $\text{pH } 7.2$, with osmolalities as in Fig. 1. Using these solutions, the Cl^- equilibrium potential (E_{Cl^-}), after correction for ionic activities⁶, was shifted from 0 mV , as was the case for solutions used in Fig. 1b, to $+54 \text{ mV}$ in Fig. 3b ($E_{\text{glutamate}} = +54 \text{ mV}$; $E_{\text{Ca}^{2+}} > +82 \text{ mV}$), whereas E_{H^+} was shifted from $+100 \text{ mV}$ to 0 mV and the equilibrium potential for free divalent cations was shifted from $+5 \text{ mV}$ to -9 mV (compare solutions in Figs 1 and 3).

with a permeability to anions. As mature guard cells are uncoupled from neighbouring epidermal cells²³, the observed $[Ca^{2+}]_{\text{cyt}}$ -induced depolarizations to about +10 mV may take place in guard cells embedded in their natural environment. Such depolarizations have been estimated to be sufficient to drive physiological rates of K^+ efflux through $I_{K^+,out}$ channels^{4,5}. Ca^{2+} can therefore trigger release of both K^+ and anions²⁰, providing a mechanism for the Ca^{2+} requisite for closure of stomatal pores¹⁰⁻¹².

The close correlations between $[Ca^{2+}]_{\text{cyt}}$ -dependent ion channels and the effects of Ca^{2+} on stomata suggest that the Ca^{2+} -dependent mechanisms reported here may provide a molecular basis for the regulation of ion transport during stomatal opening and closing. $I_{K^+,out}$ channels²⁴⁻²⁷ and, more recently, $I_{K^+,in}$ channels²⁷ with similar properties to those described in guard cells, have been characterized in other higher plant cells (for review, see refs 28, 29). Studies of effects of cytosolic Ca^{2+} on ion channels in these cells may increase our general understanding of Ca^{2+} -dependent osmoregulation in higher plant cells. □

Received 30 November 1988; accepted 14 February 1989.

- Hepler, P. K. & Wayne, R. A. *Rev. Pl. Physiol.* **36**, 397-439 (1985).
- Raschke, K. in *Encyclopedia of Plant Physiology* Vol. 7 (eds Haupt, W. & Feinlieb, M. E.) 384-441 (Springer, Berlin, 1979).
- Outlaw, W. H. *Physiologia Pl.* **59**, 302-311 (1983).
- Schroeder, J. I., Hedrich, R. & Fernandez, J. M. *Nature* **312**, 361-362 (1984).
- Schroeder, J. I., Raschke, K. & Neher, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4108-4112 (1987).
- Schroeder, J. I. *J. gen. Physiol.* **92**, 667-683 (1988).
- Blatt, M. R. *J. Membrane Biol.* **102**, 235-246 (1988).
- Assmann, S. M., Simoncini, L. & Schroeder, J. I. *Nature* **318**, 285-287 (1985).
- Shimazaki, K., Iino, M. & Zeiger, E. *Nature* **319**, 324-326 (1986).
- De Silva, D. L. R., Cox, R. C., Hetherington, A. M. & Mansfield, T. A. *New Phytol.* **101**, 555-563 (1985).
- MacRobbie, E. A. C. in *Molecular and Cellular Aspects of Ca^{2+} in Plant Development* (eds Trewavas, A. J. & Marmé, D.) 383-384 (Plenum, New York, 1986).
- Schwartz, A., Ilan, N. & Grantz, D. A. *Pl. Physiol.* **87**, 583-587 (1988).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
- Williamson, R. E. & Ashley, C. C. *Nature* **296**, 647-651 (1982).
- Wayne, R. & Hepler, P. K. *Pl. Physiol.* **77**, 8-11 (1985).
- Brownlee, C. & Wood, J. W. *Nature* **320**, 624-626 (1986).
- Hepler, P. K. & Callahan, D. A. *J. Cell Biol.* **105**, 2137-2143 (1987).
- Miller, A. J. & Sanders, D. *Nature* **326**, 397-400 (1987).
- Bush, D. S. & Jones, R. L. *Cell Calcium* **8**, 455-472 (1987).
- MacRobbie, E. A. C. *J. exp. Bot.* **32**, 563-572 (1981).
- Meech, R. W. *J. Physiol.* **237**, 259-277 (1974).
- Marty, A. *Trends Neurosci.* **6**, 262-265 (1983).
- Palevitz, B. A. & Hepler, P. K. *Planta* **164**, 473-479 (1985).
- Iijima, T. & Hagiwara, S. *J. Membrane Biol.* **100**, 73-81 (1987).
- Schauf, C. L. & Wilson, K. L. *Pl. Physiol.* **85**, 413-418 (1987).
- Moran, N. et al. *Pl. Physiol.* **88**, 643-648 (1988).
- Bush, D. S., Hedrich, R., Schroeder, J. I. & Jones, R. L. *Planta* **176**, 368-377 (1988).
- Tazawa, M., Shimmen, T. & Mimura, T. A. *Rev. Pl. Physiol.* **38**, 95-117 (1987).
- Hedrich, R. & Schroeder, J. I. A. *Rev. Pl. Physiol.* **40**, 539-569 (1989).
- Tsien, R. Y. *Biochemistry* **19**, 2396-2404 (1980).

ACKNOWLEDGEMENTS. We thank G. A. Menendez for technical assistance, Dr E. Tobin for use of plant growth chambers, Dr J. Vergara for use of Ca^{2+} electrodes, and Drs F. Bezanilla, J. Umbach, L. Byerly, E. Neher and D. Kalman for reading the manuscript. This research was supported by a grant from NIH (S.H.). J.I.S. is an Alexander von Humboldt Fellow.

Separate lineages of T cells expressing the $\alpha\beta$ and $\gamma\delta$ receptors

Astar Winoto & David Baltimore

Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

THE T-cell antigen receptor is a heterodimer molecule composed of either $\alpha\beta$ or $\gamma\delta$ chains. The $\alpha\beta$ receptor molecules are expressed mainly in $CD4^+CD8^-$ and $CD4^-CD8^+$ T cells (helper and killer T cells respectively), whereas the $\gamma\delta$ receptor molecules are expressed mainly in $CD4^-CD8^-$ T cells¹⁻³. $CD4^+CD8^-$ and $CD4^-CD8^+$ T cells arise from a class of $CD4^-CD8^-$ T cells during thymus development⁴, raising the question of whether cells

rearranging their $\gamma\delta$ receptors later give rise to $\alpha\beta$ T cells by further rearrangements of their receptor genes, or whether rearrangements and expression of the receptor genes occur in separate lineages. The δ -chain gene is located between the $V\alpha$ (variable) and $J\alpha$ (joining) gene segments^{5,6} and when the rearrangements allowing α - and β -receptors occur, the DNA between these segments is deleted as small circles which can be isolated from developing thymocytes^{7,8}. The rearrangement status of the δ -chain gene in the α -circles can therefore be investigated to see whether α -chain and δ -chain expression occur in parallel lineages or sequentially within a lineage. We find that the δ -chain gene in the T-cell receptor α -circles has a germline configuration, indicating that $\alpha\beta$ and $\gamma\delta$ T cells are distinct lineages.

To analyse the δ -chain gene rearrangement status in $\alpha\beta$ -producing T cells, two independent circle libraries were made from the thymuses of AKR mice. The circular DNA isolated was extensively purified (see legend to Table 1), cut with *EcoRI* and cloned into the λ vectors. We obtained 5×10^4 recombinants and 5×10^3 clones for the first and the second libraries respectively.

TABLE 1 Number of clones from circle libraries that hybridized to specific probes

Probes	Library 1	Library 2
<i>J\alpha</i> 26	0	0
<i>J\alpha</i> 27	>300	>100
<i>J\alpha</i> 19	55	ND
<i>J\alpha</i> 57	45	ND
A	>500	ND
B	>500	48
C	>500	46
A ⁺ B ⁺ C ⁺	>500	—
B ⁺ C ⁺	—	46
A ⁺ C ⁺	2	ND
B ⁺ C ⁺	ND	2

Several *J\alpha* oligonucleotides (*J\alpha*19, *J\alpha*57, *J\alpha*26 and *J\alpha*27)⁹ were used as probes and hybridized to the circle libraries. The *J\alpha*26 gene segment is the most $C\alpha$ -proximal, located 3 kb 5' of $C\alpha$. The *J\alpha*19 and *J\alpha*57 gene segments are located on one *EcoRI* fragment at 41 kb and 38 kb 5' of $C\alpha$, respectively. The *J\alpha*27 gene segment is located 60 kb from $C\alpha$ (ref. 9). The oligonucleotide sequences are: *J\alpha*19, ACCAGCTGAAGGTTAT; *J\alpha*57, AACCTTACAAGTGCAAC; *J\alpha*26, ACCCAAGTGACGGTAAT; *J\alpha*27, GCTGACAGTGAGCTCAT. Prehybridization was at 65 °C for 5 h in 6×SSC, 5×Denhardt's solution, 0.1% SDS and 250 $\mu\text{g ml}^{-1}$ of denatured salmon sperm DNA. Hybridization was at 37 °C overnight in the above-mentioned buffer containing oligonucleotide probe. Washing was at 37 °C for 2 h with several changes. Circle libraries were prepared essentially as described in Okazaki et al.⁸ Briefly, thymuses from 50 4–5-week-old mice were taken and washed in PBS buffer. The thymocytes were then resuspended in 20 ml nuclear homogenization buffer (0.3 M sucrose, 20 mM Tris-HCl, pH 7.6, 10 mM KCl, 0.5 mM EDTA, 3 mM CaCl₂) and homogenized with a type B pestle in a Dounce homogenizer for <10 strokes. NP40 was added to a final concentration of 0.5% and cells were homogenized again for 7 strokes, or until nuclear breakage was complete. The nuclei were then collected by centrifugation at 2,000 r.p.m. for 5 min and washed 3 times with the same homogenization buffer (30 ml) and 0.5% NP40. The final nuclear pellet was resuspended in 20 ml homogenization buffer, to which 800 μl 0.5 M EDTA, 400 μl Triton X-100 and 400 μl 20% Sarkosyl were added, and the mixture incubated for an additional 15 min on ice. The supernatant obtained after 30 min centrifugation in the SW41 rotor at 40,000 r.p.m. was poured off carefully and phenol-extracted. After extensive dialysis against 10 mM Tris-HCl, pH 7.5, 1 mM EDTA and 100 mM NaCl, the resulting material was banded in a CsCl/ethidium bromide gradient in the presence of 50 μg intact pUCPvull plasmid DNA as a carrier (pUCPvull is pUC18 plasmid with the small *PvuII* fragment removed, hence the plasmid does not contain any polylinker sequences and cannot be cloned in an *EcoRI*-cloning vector). Supercoiled DNA was collected and banded one more time. The supercoiled DNA was then extracted with butanol and ethanol-precipitated. Any residual linear DNA was digested with an ATP-dependent DNase I, leaving the supercoiled DNA intact. For cloning, the final preparation of DNA was cut with *EcoRI* and packaged into a $\lambda\text{gt}10$ vector or λZap vector (Stratagene).

ND, Not determined.

To show that the libraries consisted mainly of clones from circular DNA molecules, four $J\alpha$ oligonucleotide probes specific for the non-conserved portion of the $J\alpha$ gene and located at different distances from $C\alpha$ (constant) were used in a hybridization analysis. If the library derived from circles formed during proper rearrangements, the most $C\alpha$ -proximal $J\alpha$ segment should be absent from the library, and representation should ideally increase with 5' distance from $C\alpha$. As predicted, the probe for the most $C\alpha$ -proximal gene segment, $J\alpha 26$ (ref. 9), did not hybridize to any library clone (Table 1). By contrast, the $J\alpha 27$ oligonucleotide hybridized to several hundred clones in both libraries. Indeed, the unusually high frequency of $J\alpha 27$ -positive clones indicated that we had cloned circular DNA containing an over-representation of the T cell receptor $J\alpha$ region. Because we used *EcoRI* as the restriction enzyme to cut the circular DNA during the cloning process, hybridization with the probes for the regions $J\alpha 19$ and $J\alpha 57$ located within a single *EcoRI* restriction fragment was particularly informative in assessing the status of the libraries. In the first circle library, 55 clones hybridized to the $J\alpha 19$ probe, whereas 45 hybridized to $J\alpha 57$ (Table 1). All 45 $J\alpha 57^+$ clones were also $J\alpha 19^+$, indicating that they contained the *EcoRI* fragment of the α -chain circle in the germline configuration. Because the $J\alpha 19$ gene segment is located 5' to the $J\alpha 57$ gene segment, the rest of the $J\alpha 19^+$ clones, which were $J\alpha 57^-$, presumably contained the α -chain circles from a rearrangement between the two $J\alpha$ gene segments. The absence of $J\alpha 19^-/J\alpha 57^+$ clones indicated once again that there was no chromosomal DNA contaminant in the library.

To obtain probes for the mouse δ -chain locus, which is located about 75 kilobases (kb) 5' of the $C\alpha$ gene segment^{5,6}, we used a restriction fragment derived from the most 5' end of the cosmid clone TA28.1 (ref. 9) to isolate further cosmids (Fig. 1). Probes were derived that would hybridize to regions 5' to the $D\delta 2$ gene segment (probe A), within the $D\delta 2$ - $J\delta 1$ intron (probe B) and 3' of the $J\delta 1$ gene segment (probe C) (Fig. 1). Because all three probes are contained in a single 7.4-kb *EcoRI* restriction fragment, we anticipated that clones from the α -chain circles containing the germline configuration of the δ -chain would hybridize to all three probes, whereas clones containing rearranged δ segments would hybridize to a subset of the probes, except in the case of a rearrangement to the $J\delta 2$ gene segment, which would give no fragment.

Each of the probes we tested hybridized to >500 clones in the larger library and to 48 clones in the smaller. The number of $J\delta 1$ -positive clones was in agreement with what would be expected from a circle library with many $J\alpha$ positives, because a $J\delta$ probe will detect all the α -circles and a single $J\alpha$ probe will only detect a proportion of the α -circles. (Okazaki and Sakano¹⁰ have earlier reported analysing a circle library in which they detected only 5 $J\delta$ positives and >200 α -positives with a single $J\alpha$ probe. This apparent conflict with our results may reflect the loss of $J\delta 1$ clones during the amplification step used in the construction of their library. Our libraries were not amplified at the time of the analysis). All clones that hybridized

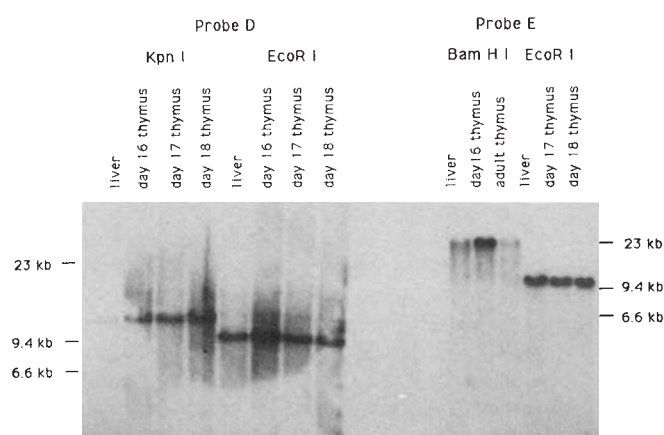


FIG. 2 Southern genomic blots of mouse liver DNA and various mouse thymus DNAs, using the probes D and E depicted in Fig. 1. Probes D and E hybridized to thymus DNAs from days 16, 17 and 18 of gestation, and from adult mouse. These DNAs were digested with several different restriction enzymes spanning the $C\delta$ - $J\alpha 1$ region (*EcoRI* and *KpnI* for probe D, and *EcoRI* and *BamHI* for probe E, see Fig. 1). This allowed us to detect any moderately frequent rearrangement into the region between the $C\delta$ and $J\alpha 1$ gene segments.

to probe C also hybridized to probes A and B (Table 1). There were also a few clones scoring A^+C^- (these were isolated in early screening and were not tested with the B probe) and B^+C^- (not tested with the A probe). We isolated several $A^+B^+C^+$ clones whose restriction fragments were indistinguishable from the germline structure presented by cosmid clone TA9. Although this result strongly implies that the α -circles have a germline δ -chain gene configuration, there are two types of circle arising from δ -chain gene rearrangements that could also score as positives: rearrangements to $J\delta 1$ and $VD\delta 2$ rearrangements would give $A^+B^+C^-$ ($VJ\delta 1$ or $VD\delta 1J\delta 1$ circles), $A^-B^+C^-$ ($D\delta 2J\delta 1$ or $VD\delta 2J\delta 1$ circles) or $A^+B^-C^-$ ($VD\delta 2$ or $VD\delta 2J\delta 1$ circles) and rearrangements to the $J\delta 2$ gene segment would give $A^+B^+C^+$ ($D\delta 1J\delta 2$ or $VJ\delta 2$ circles) or $A^-B^+C^+$ ($D\delta 2J\delta 2$ or $VD\delta 2J\delta 2$ circles). The small number of C^- cases in our library (Table 1), however, in conjunction with the fact that the number of rearrangements to $J\delta 2$ was at most 1/10 of those to $J\delta 1$ (ref. 6), indicated that most $A^+B^+C^+$ clones were derived from α -circles. We conclude that the δ -chain gene is in the germline configuration in the circle products of the α -chain gene rearrangement.

To determine whether a rearrangement analogous to the T-early- α rearrangement—which takes place between a region 5' to the $D\delta 2$ gene and a $\psi J\alpha$ region several kb upstream of the most 5' of the $J\alpha$ gene segments of the human α/δ gene locus^{11,12} also occurs in the mouse α/δ locus, we performed Southern genomic blots of mouse thymus DNA using probes lying immediately 5' of the $J\alpha 1$ gene segment (probes D and E in

FIG. 1 Schematic diagram of the T-cell antigen receptor α -chain and δ -chain gene locus. TA3 and TA9 were obtained using a restriction fragment from the most 5' end of the TA28.1 cosmid clone⁹ to screen an AKR cosmid library constructed in the pTL5 vector¹⁴. A plus sign indicates the presence of additional restriction enzyme sites. Probes A, B and C were derived from a subclone containing the 7.4-kb *EcoRI* fragment of the cosmid clone TA9. Probe A was a 1-kb *XbaI*/*PvuII* fragment, probe B was a 360-base pair *HincII*/*PvuII* fragment and probe C was a 2-kb *SacI* fragment. Probes D and E were derived from the cosmid clone TA28.1. Probe D was a 2-kb *KpnI*/*HindIII* fragment and probe E was a 1.2-kb *EcoRI*/*BglII* fragment.

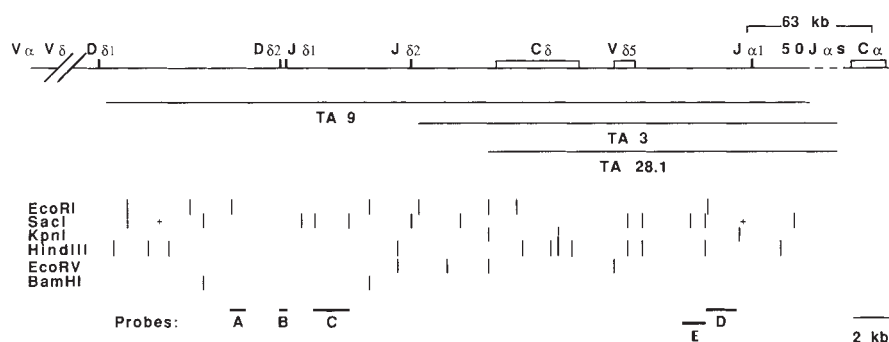


Fig. 1). Probes D and E, however, both detected only the germline bands and no rearranged fragments, even after the blots were exposed for a long time (Fig. 2). In contrast to the human T-early- α rearrangements, which were readily detectable in a Southern genomic blot of human thymus DNA, we found no analogous rearrangement in the mouse and so conclude that the majority of the circle DNAs we analysed came from the α -chain rearrangement products.

Our results help to clarify the scheme of T-cell differentiation. It has previously been suggested that progenitor thymocytes initially attempt to rearrange the genes for γ - and δ -chains; when they fail to produce functional $\gamma\delta$ receptors, the cells then rearrange their α - and β -chain genes¹³. This model is inconsistent with our analysis of the δ -chain gene rearrangement status in the α -circles, which indicates that a majority of α -producing T cells have never rearranged their δ -chain locus. Thus T cells expressing $\alpha\beta$ and $\gamma\delta$ receptors are not in the same developmental lineage. The split of $\alpha\beta$ - and $\gamma\delta$ -producing T-cell lineages could occur after the initial γ -chain gene rearrangements, known to start at day 14 of gestation. Cells with a productive γ -chain gene rearrangement would go on to rearrange their δ -chain gene, whereas cells with a non-productive γ -chain gene rearrangement would then differentiate to the $\alpha\beta$

lineage. Alternatively and perhaps more probably, the T-cell lineage decision would be made independently of the rearrangement process, with one subset of T cells committing to α -chain rearrangement and another subset to δ -chain rearrangement. Rearrangements at the γ -chain gene and the *DJ* of the β -chain gene might then happen in both lineages because of a lack of precise control over these particular events. □

Received 14 November 1988; accepted 20 February 1989.

1. Kronenberg, M., Siu, G., Hood, L. E. & Shastri, N. A. *Rev. Immun.* **4**, 529–591 (1986).
2. Davis, M. M. & Bjorkman, P. J. *Nature* **334**, 395–402 (1988).
3. Toyonaga, B. & Mak, T. W. A. *Rev. Immun.* **5**, 585–620 (1987).
4. von Boehmer, H. A. *Rev. Immun.* **6**, 309–326 (1988).
5. Chien, Y. et al. *Nature* **330**, 722–727 (1987).
6. Chien, Y., Iwashima, M., Kaplan, K. B., Elliott, J. F. & Davis, M. M. *Nature* **328**, 677–682 (1987).
7. Fujimoto, S. & Yamagishi, H. *Nature* **327**, 242–243 (1987).
8. Okazaki, K., Davis, D. D. & Sakano, H. *Cell* **49**, 477–485 (1987).
9. Winoto, A., Mjolsness, S. & Hood, L. E. *Nature* **316**, 832–836 (1985).
10. Okazaki, K. & Sakano, H. *EMBO J.* **7**, 1669–1674 (1988).
11. de Villartay, J. P. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8608–8612 (1987).
12. de Villartay, J. P., Hockett, R. D., Coran, D., Korsmeyer, S. J. & Cohen, D. *Nature* **335**, 170–174 (1988).
13. Pardoll, D. M. et al. *Nature* **326**, 79–81 (1987).
14. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. *Cell* **28**, 489–498 (1982).

ACKNOWLEDGEMENTS. We thank Dr Hee Sup Shin for the use of his cosmid library and Drs M. Lenardo, J. Pierce and S. Smale for critical reading of the manuscript. A. W. is a postdoctoral fellow of The Jane Coffin Childs Memorial Fund for Medical Research. This research was supported by the NIH.

Distal-less* encodes a homoeodomain protein required for limb development in *Drosophila

Stephen M. Cohen*, Günter Brönner*, Frank Küttner*, Gerd Jürgens* & Herbert Jäckle*

Max-Planck-Institut für Entwicklungsbiologie, Spemannstr. 35, 7400 Tübingen, FRG

THE spatial organization of the *Drosophila* embryo depends on the activity of three axial pattern-forming systems. In addition to the anterior-posterior and dorsal-ventral systems that organize the segmented body plan¹, a proximal-distal pattern-forming system is required to provide positional information for the developing limbs. The development of both the larval and adult limbs depends directly on the activity of the *Distal-less* gene^{2,3}. Genetic analysis has shown that *Distal-less* functions as a developmental switch that is required to promote the development of limb structures above the evolutionary ground-state of body wall. Here we provide genetic evidence that indicates a graded requirement for *Distal-less* activity during limb development. Reduction of this activity has a global effect on pattern formation in the limb. The molecular structure of the *Distal-less* locus indicates that the gene encodes a homoeodomain-containing protein which is therefore likely to specify limb development through differential regulation of subordinate genes.

The normal development of *Drosophila* limbs depends on the activity of the *Distal-less* (*Dll*) gene. Several mutations that reduce, but do not completely eliminate, *Dll* activity permit survival to adulthood but cause developmental abnormalities in the adult limbs³. Examination of the effects of these alleles on the adult leg shows a progressive pattern of developmental abnormalities that reflects the spatial requirement for *Dll* activity during limb development (Fig. 1). The range of severity of phenotypes caused by different alleles indicates that there is a

graded requirement for *Dll* activity in the developing leg. The most striking feature of these phenotypes is that progressively stronger alleles do not cause deletion of discrete structures in a simple distal to proximal series. Rather, structural elements are lost concomitantly from broad regions of the pattern so that the different regions of the limb are still present but are reduced in size and fused together (Fig. 1*b–d*). These results indicate that reducing *Dll* activity alters the pattern of the developing limb as a whole and suggest that the gene may therefore be responsible for organizing this pattern.

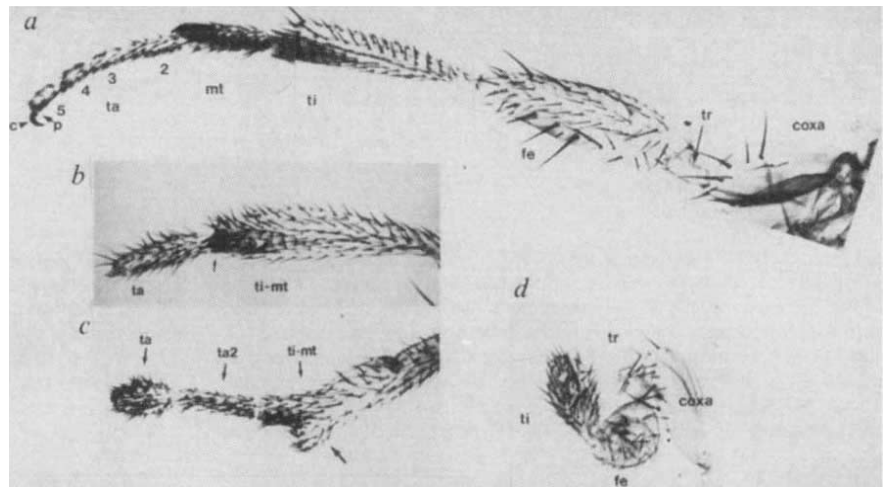
The formal genetic requirement for a graded activity of the gene does not in itself imply that the product need be distributed in a concentration gradient within the developing limb primordium. The ability of individual cells to develop into limb structures, however, depends directly on their own ability to express *Dll*². In this context, the graded requirement for *Dll* expression in individual cells suggests that the concentration of the *Dll* gene product is directly involved in specifying positional information along the proximal-distal axis of the limb.

To assess this prediction we undertook a molecular analysis of the *Dll* locus. We cloned the region by chromosomal walking and mapped the molecular lesions associated with eight alleles (Fig. 2*a*). These lesions are evenly distributed through about 38 kilobases (kb) of DNA in which only a single transcription unit, encoding transcripts of ~2.5 and ~3.4 kb, can be detected (Fig. 2*b–d*). Non-contiguous restriction fragments distributed over ~25 kb of the genomic DNA detect the same pair of transcripts. As these transcripts are rare (<10% the level of *engrailed*) we have been unable to isolate full-length complementary DNA clones and can only infer that primary transcripts derived from this relatively large region of genomic DNA are spliced to produce the messenger RNAs detected by the various probes (Fig. 2*e*).

The activity of *Dll* is known to be required during embryogenesis², as amorphic embryos fail to develop larval limbs^{2,3}. The precise time during embryogenesis at which the larval and imaginal limb primordia become determined however, is unknown. The vestigial larval limbs are peripheral sense organs² which develop during late embryogenesis, following germ band retraction⁴. The putative *Dll* transcripts are first detectable at low levels in post-blastoderm embryos and, as expected, at increased levels during late embryogenesis. Expression of the gene is also required in the developing limb imaginal disk primordia². The transcripts are expressed during the first larval

* Present address: Institut für Genetik und Mikrobiologie, Universität München, Maria Ward Str. 1a, 8000 München 19, FRG.

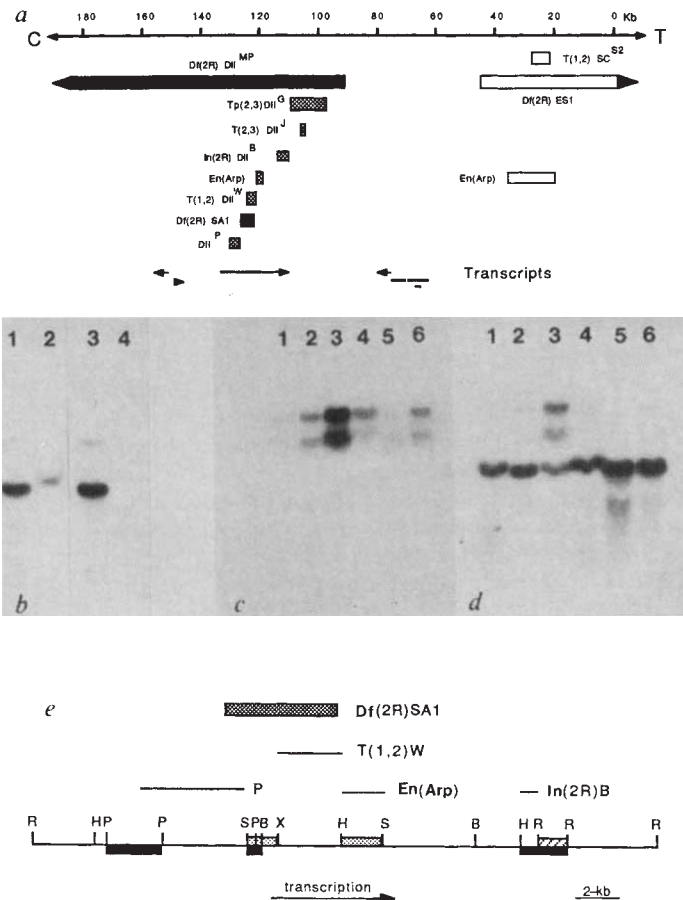
FIG. 1 Deletions of pattern elements in the legs of partial loss-of-function *Dll* mutants. *a*, A wild-type male foreleg. Coxa is the most proximal leg segment and develops independently of *Dll* activity². The more distal segments of the appendage are the limb proper²³. Abbreviations are: tr, trochanter; fe, femur; ti, tibia; mt, metatarsus (also called basitarsus); ta, tarsal segments 2–5; c, claw; p, pulvillus. In the weakest alleles (P and M, for example) only minor reductions in the tarsal region can be detected (not shown). Alleles of moderate severity (IB, 6, 7, 8) cause more extreme reduction of tarsal structures and fusion between tarsal and tibial elements. *b*, Foreleg of a male homozygous for *Dll*^{IB}. The metatarsus and tibia are fused. The sex comb of the metatarsus (arrow) and the transverse rows of bristles normally found on the tibia are now found on a common segment (ti-mt), and are both reduced. A single nondescript tarsal segment remains distally. *c*, The leg of a *Dll*^T homozygote. Note the fused metatarsus and tibia (ti-mt). *Dll*^T homozygotes typically have normal second tarsal segment (ta2) and a bulbous distal tarsal blob (ta). *d*, Foreleg of a male of the most extreme viable genotypic combination *Dll*³/*Dll*^{SA1}. Note the greatly reduced femur and tibia associated with normal trochanter and coxa. All independent tarsal segments seem to be deleted. In rare cases of a milder genotype *Dll*³/*Dll*³ an occasional sex comb tooth can be detected on the most distal segment, which suggests that this structure may be a fused metatarsus-tibia. Analogous defects also observed in the other segmented appendages of the *Drosophila* adult: the antenna, the maxilla, the labium and



the labrum. Detailed morphological descriptions of the *Dll* hypomorphic phenotypes in these appendages will be presented elsewhere (S.M.C. and G.J., manuscript in preparation).

METHODS. Adult cuticle preparations were as previously described². Some of the *Distal-less* alleles used here have been described previously under the name *Brista*³. We have retained the previous allele designations. The origins of the other alleles will be presented elsewhere (S.M.C. and G.J., manuscript in preparation).

FIG. 2 Molecular organization of the *Dll* locus at cytogenetic map position 60E5.6 on chromosome 2R. *a*, The molecular locations of chromosomal rearrangements associated with *Dll* alleles are indicated by filled boxes. The size of the black boxes indicates the extent of the deletions in the alleles MP and SA1. The size of the hatched boxes indicates the uncertainty in the exact location of the breakpoints of the translocations W and J, the transposition G, the insertions P and *En(Arp)* and the inversion B. Open boxes indicate the locations of rearrangements that genetically define a distal limit for the *Dll* locus. *En(Arp)* was generated in a hybrid dysgenic screen (P. Adler) and contains two independent insertions in the region: an insert of 8-kb of DNA associated with the mutation and a P-element near coordinate 30 which lies outside the gene (defined by the breakpoint of *Df(2R)ES1* which is *Dll*⁺). C and T indicate direction of centromere and telomere. The location and orientation (where known) of transcription units within the interval between 60- and 160-kb are indicated. *b*, Reduced stringency hybridization to genomic DNA from *D. melanogaster* (lanes 1, 3) and *D. virilis* (lanes 2, 4) was used to locate coding sequences²⁴. The two species are diverged by ~60 million years²⁵. Lanes 1 and 2 were probed with the 1.6-kb *SalI*-*XbaI* fragment (stippled box in *e*) located within the DNA removed by the 5.5-kb deletion associated with the SA1 allele. Lanes 3 and 4 were probed with the 2-kb *HindIII*-*SalI* fragment flanking the deletion. Note that although both probe fragments hybridize to the same 6.6-kb genomic *SalI* fragment in the *D. melanogaster* DNA, only the 1.6-kb *XbaI*-*SalI* probe detects a conserved sequence in the *D. virilis* DNA. *c*, Transcription of the *Dll* locus. 10 µg of poly(A)⁺ RNA from 0–4 h embryos, 4–8 h embryos, 8–20 h embryos, and first, second and third instar larvae (L1, L2 and L3) in lanes 1–6, respectively, were hybridized with the 1.4-kb *EcoRI* fragment encoding the 3' region of the transcript (indicated by the diagonally striped box in panel *e*). *d*, Same filter as *c*, rehybridized with a tubulin cDNA clone. Note that the apparent concentration of the *Dll* RNAs in the 8–20 h sample under-represents the actual relative level because this lane is underloaded. *e*, Location of the proposed *Dll* exons on the physical map of the locus. Restriction fragments that hybridize strongly to the transcripts in *c* are indicated by filled boxes below the map. B, *Bam*HI; H, *Hind*III; P, *Pst*I; R, *Eco*RI; S, *Sal*I; X, *Xba*I. Within the 870-base pair (bp) *SalI*-*Bam*HI fragment both the 500-bp *SalI*-*Pst*I and the adjacent 370-bp *Pst*I-*Bam*HI subfragments hybridized to the transcript. DNA sequence analysis reveals an open reading frame spanning the *Pst*I site that has the capacity to encode part of a homeodomain. The locations of the transcribed sequences are consistent with the molecular lesions associated with the amorphic alleles. *Dll*^{MP} is a large deletion which removes the entire coding region. The small deletion SA1 removes a single exon. The deletion in the SA1 chromosome does not exist in the parental chromosome from which the mutant line was derived and is therefore likely to be the direct cause of the mutant phenotype. The fragments in which the breakpoints of other alleles are located are indicated by lines above the map. The translocation allele W and the inversion B would prevent correct splicing of the mature transcript. The insertion of ~8-kb into the ~13-kb intron in the *En(Arp)* allele also causes an amorphic phenotype. The translocation breakpoint in the allele J is located 3' to the proposed coding region.



METHODS. Molecular breakpoints of the chromosomal rearrangements were determined by *in situ* hybridization to polytene chromosomes using biotinylated probes²⁶ and by genomic Southern hybridization. The origins and genetic properties of the alleles are described in ref. 2. Reduced-stringency Southern blots were washed in 1 × SSPE at 65 °C. RNAs were fractionated on formaldehyde agarose gels, transferred to nylon and hybridized with ³²P-labelled probes generated by random priming of gel-purified restriction fragments. The orientation of transcription was determined using single-stranded RNA probes and by DNA sequence analysis of short cDNA clones from the 3' ends of the transcripts (not shown).

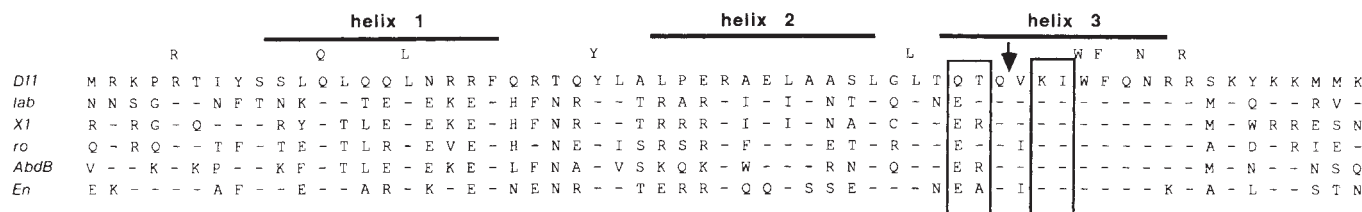


FIG. 3 Sequence comparison of the *Dll* and other homoeodomains. Dashes indicate sequence identity with *Dll*. The positions of the α -helices determined by NMR analysis of the antennapedia protein²⁷ are indicated. Boxed residues correspond to solvent-exposed residues involved in sequence-specific contact to the DNA in the bacterial helix-turn-helix proteins²⁸. Amino-acid residues conserved in all characterized *Drosophila* homoeodomains are indicated above the *Dll* sequence. The *Dll* HD is diverged from other *Drosophila* HDs with a maximum of 48% amino acid identity to *engrailed*^{29,30}, *AbdB*³¹ and

*zen2*³². In spite of lower overall sequence identity to *labial*⁹ (46%) meaningful similarity is greater in that both are interrupted by an intron between residues 44 and 45 (arrow). Other HDs shown for comparison are *rough*³³, which specifies cell identity in eye development and is like *labial*, and the *Xenopus* *XlHbox1* HD¹³, which may play a role in vertebrate limb development (indicated as X1). Note that this protein, and its newt and human homologues (not shown), are less similar to *Dll* than they are to other *Drosophila* HDs such as *labial*.

instar, at reduced levels during second instar and at higher levels during third instar. The biphasic expression profile is consistent with the requirements for *Dll* activity in development of the larval and adult appendages^{2,3}.

The DNA sequences of the genomic fragments that hybridize to the transcripts were determined. A potential protein-coding region was found that has the capacity to encode a polypeptide containing the first 44 amino acids of the homoeodomain (HD), a DNA-binding protein motif found in several developmental regulatory genes of *Drosophila*, yeast and vertebrates⁵⁻⁷. The predicted protein sequence diverges significantly from the HD consensus at the position of a potential splice donor site located in the putative DNA-binding helix, and the open reading frame terminates soon thereafter, which suggests that the recognition helix is interrupted by an intron. The remainder of the HD was located in the next more 3' DNA fragment hybridizing to the RNA. This proposed exon was located as an open reading frame following a splice acceptor site which would, if used, complete the HD (Fig. 3). Although we have no direct evidence that the ~13-kb intron is spliced in this way from the *Dll* transcript, the most highly conserved region of most HDs⁸ is located 3' to the proposed splice acceptor site. Precedent exists in the case of the *labial* gene of *Drosophila*, which is spliced at precisely the same position in the HD⁹. Several HD-containing proteins bind to DNA in a sequence-specific manner^{6,10-16}, and in some cases DNA-binding activity has been shown to reside in the homoeodomain^{6,10,16,17}. At present we have too little information about the structural basis for sequence-specific DNA-binding of *Drosophila* HD proteins to assess the significance of the rather divergent amino-acid sequence of the *Dll* HD (Fig. 3).

Recently two other HD-encoding gene products have been shown to exhibit spatial expression patterns consistent with a role for these genes in limb development. The *Xenopus* *XlHbox1* protein (and its murine homologue) are distributed in a concentration gradient along the anterior-posterior axis in the mesoderm of the developing frog and mouse forelimb buds, respectively¹⁸. The newt homologue of these genes *NvHbox1* is differentially expressed along the proximal-distal axis of the regenerating newt limb¹⁹. The patterns of expression of these gene products indicate that they play a part in development of the vertebrate limb which is different from that played by *Distal-less* in the *Drosophila* limb. The vertebrate proteins are, in fact, more closely related at the amino-acid sequence level to other *Drosophila* HDs such as *labial* than they are to the *Dll* protein (Fig. 3).

The molecular characteristics of the putative *Dll* gene product are consistent with its proposed genetic function. The graded

requirement for the activity of the *Dll* locus for development of the proximal-distal pattern of the limbs is analogous to the graded requirement for the activity of *bicoid* along the anterior-posterior axis of the *Drosophila* embryo²⁰. The graded requirement for *bicoid* activity correlates with a graded distribution of the gene product in the early embryo^{21,22}. A graded distribution of the *Dll* protein provides the simplest explanation for the genetic results, although we emphasize that this is not the only possible explanation. If such a gradient exists, the mechanism by which it is established will certainly be different from the simple diffusion mechanism that has been proposed to establish the *bicoid* gradient²¹, as a gradient of the *Dll* protein would have to be established in a field of epithelial cells, rather than in a field of naive nuclei occupying a common cytoplasm. Because the *Dll* protein contains a homoeodomain, it is tempting to speculate that the concentration of the product might, like *bicoid*, directly specify positional information through differential regulation of subordinate genes. □

Received 30 January; accepted 13 February 1989.

- Ingham, P. W. *Nature* **335**, 25-34 (1988).
- Cohen, S. M. & Jürgens, G. *EMBO J.*, submitted.
- Sunkel, C. E. & Whittle, J. R. S. *Roux's Archs dev. Biol.* **196**, 124-132 (1987).
- Campos-Ortega, J. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1985).
- Gehring, W. J. *Science* **236**, 1245-1252 (1987).
- Hall, M. N. & Johnson, A. *Science* **237**, 1007-1012 (1987).
- Balling, R., Deutsch, U. & Gruss, P. *Cell* **55**, 531-535 (1988).
- Barad, M., Jack, T., Chadwick, R. & McGinnis, W. *EMBO J.* **7**, 2151-2161 (1988).
- Mlodzik, M., Fjose, A. & Gehring, W. J. *EMBO J.* **7**, 2569-2578 (1988).
- Desplan, C., Theis, J. & O'Farrell, P. *Nature* **318**, 630-635 (1985).
- Hoey, T. & Levine, M. *Nature* **332**, 858-861 (1988).
- Driever, W. & Nüsslein-Volhard, C. *Nature* **337**, 138-143 (1988).
- Cho, K. W. Y. et al. *EMBO J.* **7**, 2139-2149 (1988).
- Müller, M. et al. *EMBO J.* **7**, 4299-4304 (1988).
- Beachy, P. A., Krasnow, M. A., Gavis, E. R. & Hogness, D. S. *Cell* **55**, 1069-1081 (1988).
- Hoey, T., Warrior, R., Manak, J. & Levine, M. *Molec. cell. Biol.* **8**, 4598-4602 (1988).
- Desplan, C., Theis, J. & O'Farrell, P. *Cell* **54**, 1081-1090 (1988).
- Oliver, G., Wright, C. V. E., Hardwicke, J. & DeRobertis, E. M. *Cell* **55**, 1017-1024 (1988).
- Savard, P., Gates, P. B. & Brookes, J. P. *EMBO J.* **7**, 4275-4282 (1988).
- Frohnhöfer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120-125 (1986).
- Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 83-93 (1988).
- Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 95-104 (1988).
- Snodgrass, R. E. *Principles of Insect Morphogenesis* (McGraw-Hill, London, 1935).
- Page, D. et al. *Cell* **51**, 1091-1104 (1987).
- Beverley, S. M. & Wilson, A. C. *J. molec. Evol.* **21**, 1-13 (1984).
- Langer-Safer, P. R., Levine, M. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4381-4385 (1982).
- Otting, G. et al. *EMBO J.* **7**, 4305-4309 (1988).
- Pabo, C. O. & Sauer, R. T. A. *Rev. Biochem.* **53**, 293-321 (1984).
- Poole, S., Kauvar, L., Drees, B. & Kornberg, T. *Cell* **40**, 37-43 (1985).
- Fjose, A., McGinnis, W. & Gehring, W. J. *Nature* **313**, 284-289 (1985).
- Weinzierl, R., Axton, M., Ghysen, A. & Akam, M. *Genes Dev.* **1**, 386-397 (1987).
- Rushlow, C., Doyle, H., Hoey, T. & Levine, M. *Genes Dev.* **1**, 1268-1297 (1987).
- Saint, R., Kalionis, B., Lockett, T. J. & Elizur, A. *Nature* **334**, 151-154 (1988).

ACKNOWLEDGEMENTS. We thank our colleagues for their support and Claudio Sunkel for providing the *Dll*³ allele. This work was supported by the Medical Research Council of Canada (S.C.) and the Deutsche Forschungsgemeinschaft (Leibniz-Program) (H.J.).

Increased γ -globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor

David I. K. Martin, Shih-Feng Tsai & Stuart H. Orkin

Division of Hematology-Oncology, Children's Hospital, and the Dana Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

IN man, a shift from γ - to β -globin gene expression in erythroblasts underlies a switch from fetal to adult haemoglobin during development¹. In hereditary persistence of fetal haemoglobin (HPFH), inappropriately high γ -globin expression in adult life is associated with deletions in the β -globin cluster or with single-base changes upstream of the γ -globin genes. To account for enhanced γ -gene expression in HPFH of the non-deletion type, we tested the nuclear proteins of human erythroleukaemia cells² that bind γ -promoter sequences *in vitro* by correlating specific mutations in their binding sites with promoter activity. An erythroid-specific factor (GF-1) binds as a single molecule to the -195 to -170 region and contacts two TATCT(AGATA) motifs, but not the conserved octamer (ATGCAAAT)³⁻⁸ that separates them. We observe that a single change (at -175, T $\rightarrow$ C) found in HPFH^{9,10} leads to increased promoter activity only in erythroid cells. This effect is mediated by GF-1, the human counterpart of the chicken erythroid factor¹¹ Eryf 1. The form of HPFH we studied here is an inherited disorder which can be ascribed to the action of a cell-specific DNA-binding factor on a mutant promoter.

Using γ -globin-producing human erythroleukaemia (K562) cells, we sought nuclear factors which bind *in vitro* to a -257 to -140 γ -promoter fragment that contains the octamer^{12,13}, activates a linked β -globin gene on transfer into K562 cells¹⁴, and is the site of all but one of the single-base changes known to be associated with HPFH¹. Two types of binding activity were detected (Fig. 1a): one produces gel-retardation complex 2A, is found in all cell lines and is identical to the ubiquitous octamer-binding factor NF-A1/OTF-1; the other gives rise to complexes 1A/1B, is produced only by erythroid cell extracts and is termed GF-1. Binding specificities were distinguished by competition for gel-retardation complexes (Fig. 1b). The core octamer sequence, but not a heterologous sequence or DNA containing the -175 T $\rightarrow$ C HPFH substitution, inhibited formation of complex 2A. Conversely, complexes 1A/1B were competed out by DNA with the -175 T $\rightarrow$ C substitution, but not by the core octamer or a heterologous sequence. All complexes were specifically competed out by homologous DNA. Figure 2(a-e) summarizes DNase I footprinting¹⁵ and methylation interference¹⁶ assays of crude and fractionated K562 and HeLa nuclear extracts. Fractions yielding complex 2A (of K562 or HeLa origin) protect and contact nucleotides of the core octamer, consistent with its designation as NF-A1/OTF-1. Fractions containing activity-generating complex 1A/1B protect sequences extending from -195 to -170 and induce hypersensitive sites (Fig. 2a, b). Although the octamer is protected by binding of the erythroid-specific factor, DNA contact sites (Fig. 2c and d) are found only in the flanking sequences.

Several features of the binding of GF-1 to the γ -promoter merit discussion. GF-1 directly contacts repeats of five bases (TATCT or AGATA) in the footprinted region (Fig. 2e). These encompass positions -189 to -185 and -175 to -171. The similarity of DNA contact sites in these repeats suggests two regions that might bind one or two protein molecules. Although two GF-1 molecules could possibly bind to the two motifs in the -195 to -170 region, the following mutation analysis indicates that only a single molecule binds to generate complex 1A.

Complex 1B can be resolved into two species on gels (Fig. 3a): both result from binding of proteolytic fragments of GF-1 (our unpublished data). C $\rightarrow$ A substitutions at contact sites in either motif (-186 or -172) abolish one or other of the two 1B species. These mutations reduce, but do not abolish, complex 1A. A methylation-interference assay (Fig. 2d) shows complete interference of both motifs in complex 1A. If complex 1A contained two bound GF-1 molecules, either mutation (-186 or -172) would be expected to completely abolish complex 1A and also to generate a faster migrating complex, which is not observed. Furthermore, gel-retardation with probes containing single binding motifs from other globin promoters and enhancers produce complex 1A as the chief species (not shown). A double mutation at -186/-172 abolishes binding to the A γ -promoter (Fig. 3b). We conclude that a single GF-1 molecule interacts with a bipartite binding site in the γ -promoter and mutations in either motif can influence formation of the protein-DNA complex.

Evans *et al.*¹¹ recently described an erythroid-specific globin DNA-binding activity in chickens (Eryf 1) which recognizes a proposed consensus (A/T)GATA(A/G). GF-1 and the independently identified human factor NF-E1 (refs 17 and 18) seem to be the counterpart of this chicken factor. The specific competition for GF-1 binding to the γ -promoter by sequences derived from the 3'- β -globin enhancers of chicken or human origin

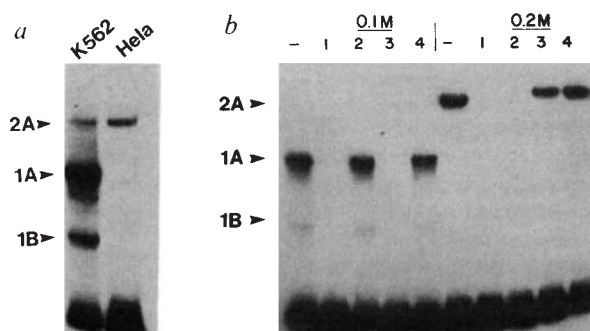


FIG. 1 Detection of DNA-binding activities in K562 and HeLa cell nuclear extracts. a, Gel-retardation assay of crude K562 and HeLa nuclear extracts incubated with a 117-base pair (bp) *HinfI/NcoI* fragment (-257 to -140) of the A γ -globin gene¹⁹ promoter. Complexes 1A/1B have been observed only with nuclear extracts of K562, HEL and MEL cells and were absent in human B- and T-lymphoid lines (EBV-line 721 and MOLT 4), monocytic U937 and THP-1, human myeloid HL60 and PLB-985, and murine myeloma cells (P3X) (not shown). b, Competition of gel-retardation complexes by double-stranded oligonucleotides. Heparin-Sepharose-fractionated K562 cell nuclear extract was used in gel-retardation assays with the *HinfI/NcoI* fragment without added competitor (-), or with addition of 20 ng (400-fold excess) double-stranded oligonucleotides (lanes 1-4). The oligonucleotides were as follows: lane 1, a 26-bp γ -promoter sequence from -195 to -169; lane 2, a 32-bp sequence containing four repeats of the ATGCAAAT motif; lane 3, a 26-bp sequence identical to that in lane 1, except for a T $\rightarrow$ C replacement at -175; and lane 4, a 39-bp sequence extending from -140 to -101 of the γ -promoter, which does not share sequences with the *HinfI/NcoI* fragment. For analysis of complexes 1A and 1B we used a 0.1 M (flow-through) heparin-Sepharose fraction, and a 0.2 M fraction for complex 2A. METHODS. Nuclear extracts were prepared as described²⁷. For preparation of K562 extract it is necessary to include leupeptin, pepstatin and aprotinin in addition to phenylmethyl sulphonyl fluoride to inhibit proteolysis. Fractionation on heparin-Sepharose was by loading crude extract in buffer D²⁷ containing 0.1 M KCl at 4°C. Step elutions with buffer D containing 0.2 M, 0.3 M, and 1.0 M KCl were carried out after collection of the flow-through (0.1 M). The 117-bp *HinfI/NcoI* fragment of the A γ -promoter¹⁹ was 5'-end-labelled and prepared by standard methods.²⁸ Gel-retardation assays^{29,30} were performed on ice in 20 μ l containing 1 μ l (0.5-5 μ g) crude or fractionated extract, 1-5 μ g poly (di-dC), and 10⁴ d.p.m. end-labelled DNA (0.2 ng) in 10 mM HEPES, pH 7.8, 50 mM potassium glutamate, 5 mM MgCl₂, 1 mM EDTA, 1 mM dithiothreitol and 5% glycerol. Incubation was for 30 min on ice. Reactions were loaded on to 0.5 $\times$ Tris-borate-EDTA/5% polyacrylamide gels and electrophoresed at 10 V per cm for 2 h at 4°C.

supports this conclusion (Fig. 3c). In contrast with our results for the γ -promoter, the two Eryf 1 binding sites in the chicken 3'- β -globin enhancer act independently¹¹. Although the motifs are comparably spaced in the two elements, their orientations are opposite in the chicken β -enhancer, but the same in the γ -promoter. GF-1 binding *in vitro* (Fig. 3c) seems to be dominant over NF-A1 (OTF-1) binding in crude extracts and excludes simultaneous occupancy by this ubiquitous factor on the promoter.

We evaluated the relevance of the binding of two factors *in vitro* with overlapping sequences in the -195 to -170 region to the enhanced expression of γ -globin in HPFH. Promoter activity was assayed in a transient expression system in which wild-type and mutant promoters (encompassing sequences from -385 to +50 of the A_γ gene¹⁹) linked to the human growth hormone (GH) gene²⁰ as a reporter were introduced into K562 cells (which synthesize γ - and ϵ -, but not β -globins), β -globin-producing mouse erythroleukaemia (MEL) cells, and non-erythroid cells. When introduced into K562 and MEL cells in supercoiled plasmid, the -175 T $\rightarrow$ C HPFH promoter directed production of 4.4 and 4.8 times as much growth hormone as the wild-type respectively ($P < 0.001$). Expression directed by the mutant promoter was not increased in two non-erythroid cell

lines, HeLa and PLB-985²¹. As the -175 T $\rightarrow$ C substitution greatly reduces binding of NF-A1/OTF-1 *in vitro*¹⁷ (Fig. 1), it has been proposed that alleviation of negative regulation by this factor might account for the HPFH phenotype of the -175 T $\rightarrow$ C mutant^{12,13}. Evidence against this model was provided by mutant promoters with other substitutions in the octamer. Expression was not enhanced by either a triple substitution (GCA $\rightarrow$ AAG at -180 to -178) or a single-base change (C $\rightarrow$ A at -179) in the octamer. *In vitro* binding to these mutant promoters by the octamer-binding factor, but not by GF-1, is drastically reduced (not shown). The role of GF-1 was investigated by mutating its DNA contact sites flanking the octamer in the context of the wild-type and -175 T $\rightarrow$ C promoters. Introduction of C $\rightarrow$ A substitutions at -186 and -172 or a single C $\rightarrow$ A change at -186 in the -175 T $\rightarrow$ C promoter reduced expression in K562 to the wild-type level. From these transient expression experiments we conclude that (1) the -175 T $\rightarrow$ C HPFH substitution increases promoter strength in erythroid cells of either fetal/embryonic (K562) or adult (MEL) type; (2) the altered expression of the -175 T $\rightarrow$ C promoter in erythroid cells is not due *per se* to reduced binding of a negatively acting octamer-binding factor; and (3) enhanced, erythroid-specific expression of the -175 T $\rightarrow$ C promoter depends on binding of GF-1.

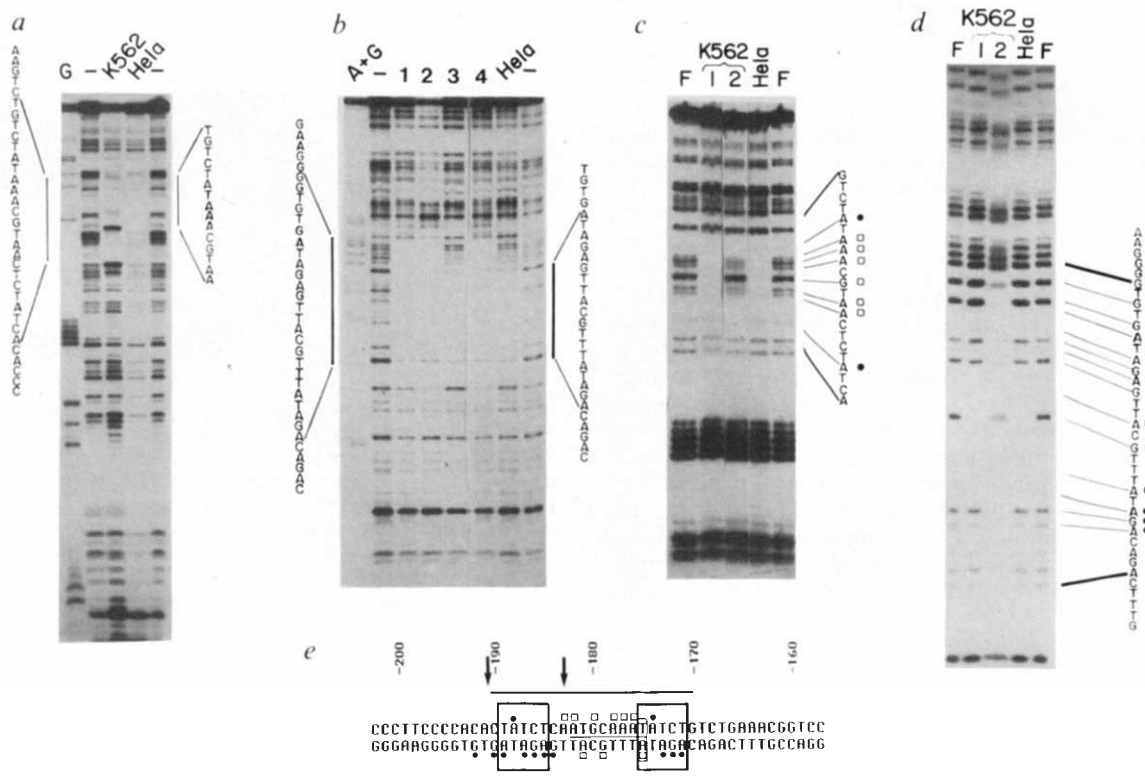


FIG. 2 DNase I footprinting and methylation interference assay of the two DNA-binding activities present in K562 nuclear extracts. *a* and *b*, DNase I footprinting of the *HinfI/NcoI* γ -globin promoter fragment by K562 and HeLa cell nuclear extracts. Coding (*a*) and non-coding (*b*) strands of the 117-bp fragment were digested with DNase I as described below. Maxam-Gilbert sequencing reactions are shown to the left in each panel. DNase digestions without added protein are denoted as (-). *a*, Crude nuclear extracts of K562 and HeLa are presented. *b*, Crude (lane 1) and heparin/Sepharose-fractionated K562 (0.1 M, 0.2 M and 1.0 M fractions, lanes 2-4) extracts and crude HeLa extract. Footprinted areas are denoted by the vertical lines at the left (K562) and right (HeLa) of each panel. For heparin/Sepharose fractionated K562 extracts, DNase I footprints obtained with 0.1 M and 1.0 M fractions are considered equivalent, as we subsequently found that the flow-through (0.1 M) fraction could be quantitatively bound and eluted from the column in the 1.0 M fraction (not shown). *c* and *d*, Methylation interference assays of complexes. G $\rightarrow$ A sequence of unbound (F) fragment DNA is shown for

reference. Interference patterns in K562 complex 2A (lane 1) and complex 1A (lane 2) and HeLa complex 2A (HeLa lane) are shown for the coding (*c*) and non-coding (*d*) strands. Open squares denote methylated bases that interfere with binding in complex 2A; closed circles denote methylated bases interfering with complex 1A. The DNase I footprints and methylation interference DNA contact sites are summarized in *e*. Horizontal lines denote the sequences on each strand protected by K562 extract (total or 0.1 M/1.0 M fractions). Duplicated and contacted TATCT (AGATA) motifs straddling the conserved octamer are boxed. Downward arrows indicate the positions of DNase-hypersensitive sites.

METHODS. DNase I footprinting was as described¹⁵. Binding was in 50 μ l containing 2.5 $\times 10^4$ d.p.m. end-labelled DNA, 1-5 μ g poly (dl-dC), and 5-50 μ g crude or fractionated nuclear extract in 10 mM HEPES, 1 mM dithiothreitol, pH 7.8, 50 mM potassium glutamate, 5 mM MgCl₂, 1 mM EDTA, 5% glycerol and 2% polyvinyl alcohol. Methylation interference assays³¹ and end-labelled probe methylated³² were as described.

Whereas enhanced expression seems to require potential GF-1 contact sites, mutations at -186 , -172 , or $-186/-172$ revealed no effect on transient expression in the context of the wild-type promoter. As γ -GH constructs have a basal promoter activity which is not cell-type specific, we also assessed the role of GF-1 binding in the -195 to -170 region on the γ -promoter function, using a stable selection assay which exhibits greater tissue-specificity²². γ -Promoter/neomycin-resistance gene constructs were introduced into K562 cells and the number of G418-resistant colonies was scored. Consistent with the transient expression data, the -175 T $\rightarrow$ C/*neo* plasmid yielded 3.1 times more G418-resistant colonies than the wild-type construct ($P < 0.01$). C $\rightarrow$ A mutations at $-186/-172$ in the presence or absence of the -175 base change reproducibly reduced the frequency of G418-resistant clones 40–50% (relative to wild-type) ($P < 0.05$). Additional regulatory elements in the -385 to $+50$ promoter fragment (outside the -195 to -170 region) thus contribute to its erythroid specificity.

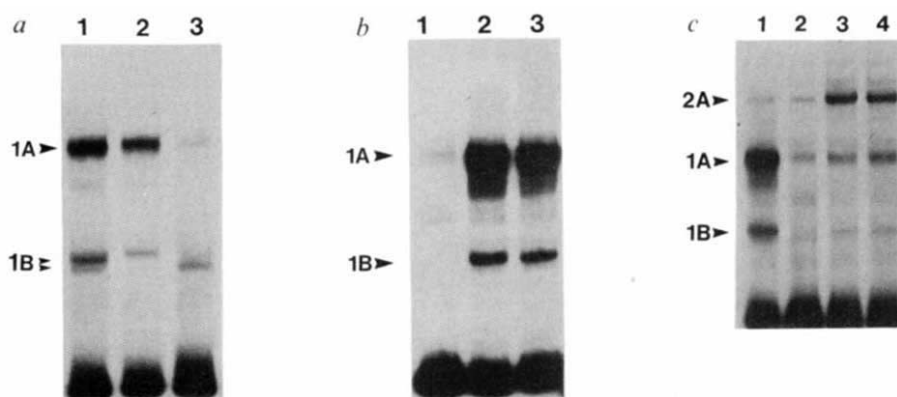
Although GF-1 is implicated in the enhanced activity of the -175 T $\rightarrow$ C mutant promoter, no major difference in its binding was apparent in the retardation assay (Fig. 3b). It has been suggested that affinity of the -175 T $\rightarrow$ C mutant promoter for

(presumably) the same erythroid-specific factor is increased 5-fold¹⁷, but we do not find this. Using either crude or fractionated nuclear extracts, we observe no difference in the affinity of GF-1 for the two promoters (Fig. 4a, b). This finding has recently been supported independently²³. Although the -175 substitution alters the proximal TATCT motif to CATCT, GF-1 still binds strongly to the proximal motif, as assayed by methylation interference (ref. 17 and our unpublished data). Disruption of the proximal motif, either alone or in combination with a substitution in the octamer (C $\rightarrow$ A at -172 and C $\rightarrow$ A at both $-172/-179$), does not mimic the effect of the -175 T $\rightarrow$ C mutation on expression. But the DNase I footprint of the mutant promoter shows reduced protection of the proximal motif, which implies an altered recognition of the bipartite site by GF-1 (Fig. 4c). Finally, we have observed that linearization of the -175 T $\rightarrow$ C promoter/GH construct before transient expression in K562 cells largely abolishes enhanced expression.

How then does the -175 T $\rightarrow$ C mutation lead to enhanced expression? Taken together, our findings suggest that the increased expression characteristic of the HPFH mutant is due to an altered interaction of the mutant promoter with GF-1. Although we find no evidence for a role for the ubiquitous

FIG. 3 Binding of GF-1 to mutant promoter DNAs

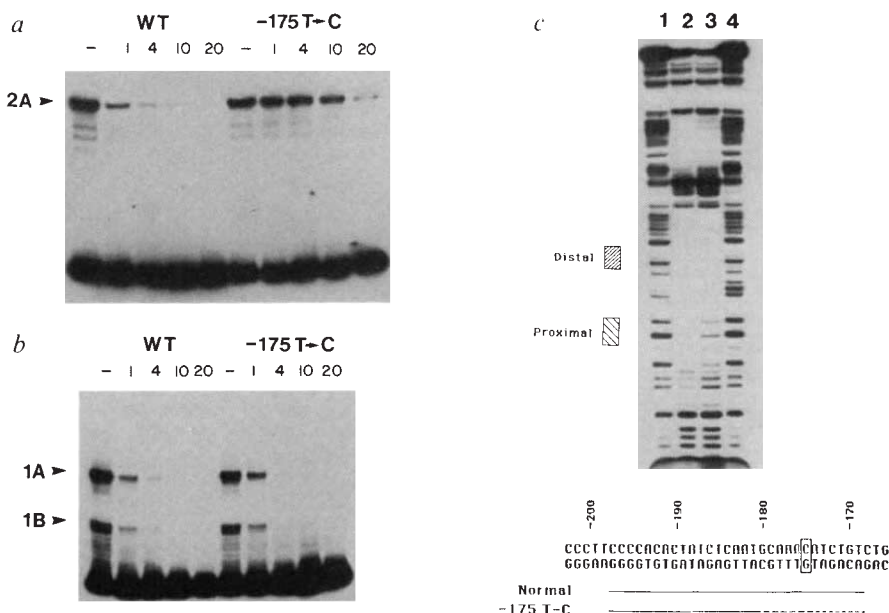
and cross-competition by human and chicken 3' β -enhancer sequences. **a**, Gel-retardation assay of K562 nuclear extract with the 117-bp *Hinf*I-*Nco*I fragment of promoters containing the normal sequence (lane 1), single C $\rightarrow$ A substitutions at -186 (lane 2) and -172 (lane 3). **b**, Gel-retardation assay of K562 nuclear extract (0.1 M heparin-Sepharose fraction) with the 117-bp *Hinf*I-*Nco*I fragment of promoters containing T $\rightarrow$ C, C $\rightarrow$ A, and C $\rightarrow$ A substitutions at -175 , -186 , -172 , respectively (lane 1), T $\rightarrow$ C at -175 (lane 2) and wild-type sequence (lane 3). A double mutant of C $\rightarrow$ A at $-186/172$ in the absence of the -175 substitution behaves similarly to the triple mutant (not shown). **c**, Complexes formed on incubation of the wild-type *Hinf*I-*Nco*I fragment with crude K562 nuclear extract (lane 1) were used to compete with ~ 100 -fold excess of double-stranded oligonucleotides: wild-type γ -promoter sequence -193 to -169 (lane 2); human 3' β -globin enhancer TGCTGGCTCCCTATCATGTCCCTATGGTGC (lane 3); chicken 3' β -enhancer



En4 (lane 4)¹¹. Note that competition for GF-1 binding (lanes 3,4) is associated with increased NF-A1/OTF-1 binding.

FIG. 4 Interaction of GF-1 with -175 T $\rightarrow$ C HPFH

DNA. In **a** and **b**, the relative affinity of K562 nuclear factors for wild-type and -175 T $\rightarrow$ C HPFH promoter sequences is compared. Competitions of formation of complex 2A (octamer-binding protein) and complexes 1A/1B (GF-1) are shown in **a** and **b**, respectively. Heparin-Sepharose-fractionated K562 extract was used in gel-retardation assays. Complex formation was competed with 0 (—), 1, 4, 10 or 20 ng unlabelled wild-type (WT) or mutant (-175 T $\rightarrow$ C) oligonucleotide (-193 to -169), as indicated. Note that the -175 T $\rightarrow$ C mutation reduces the affinity of promoter DNA for the octamer-binding factor by about 20-fold (**a**), but does not appreciably alter the affinity for GF-1 (**b**). Competition was similarly observed with larger wild-type and mutant oligonucleotides (-217 to -162) (data not shown). **c**, Altered binding of the -175 T $\rightarrow$ C mutant DNA to GF-1. DNase I footprinting was as in Fig. 2, except that GF-1 was purified by affinity chromatography, SDS-polyacrylamide gel electrophoresis, denaturation and renaturation. Purified GF-1 is a polypeptide of relative molecular mass 50,000 (S-F.T. and S.H.O., unpublished data). The non-coding strands of the normal (2) and -175 T $\rightarrow$ C mutant (3) are presented. DNase digestion without added protein for normal DNA (1) and -175 T $\rightarrow$ C DNA (4). The positions of the two TATCT motifs are indicated to the left. The extents of the footprints seen with the purified protein on the DNAs are depicted below with a dashed line depicting the reduced protection on mutant DNA. The



-175 mutant sequence is boxed. An additional footprint, present in the -215 region in the vicinity of a low-affinity GATA motif, is seen on both templates with purified factor.

octamer-binding protein in producing the HPFH phenotype, we do not dismiss the possibility that an octamer-binding factor may be involved in γ -gene regulation at some stage *in vivo*. The single base change at -175 alters the interaction of GF-1 with the promoter, rather than simply influencing their mutual affinity as assayed in gel-retardation experiments. The altered footprint of GF-1 on the HPFH mutant DNA could reflect a shift in binding from a proximal (inhibitory) site to a distal (stimulatory) site, but methylation interference demonstrates interaction with both motifs and argues against this simple model. The HPFH mutation seems to activate a region of the A_{γ} -promoter that normally contributes only a fraction of its overall activity. The effect of the -175 T→C substitution is unique in that other putative HPFH mutations (for example, at positions -202 and -196; refs 24 and 25) do not lead to striking effects on promoter strength in transient K562 or MEL cell assays (our unpublished data).

Binding motifs for GF-1 (Eryf 1, NF-E1) are widely distributed among the regulatory elements of the globin genes^{11,17,26}. Evidence presented here and elsewhere¹¹ establishes a functional role for GF-1 in the human γ -promoter and in the chicken 3' β -enhancer. This factor acts on genes expressed at different developmental stages and, as such, may participate widely in regulation of the globin locus. Our work demonstrates that GF-1 is a positive regulator of γ -gene expression and that a small change in the sequence to which it binds can result in substantially increased promoter activity. If the activity of this factor when bound to DNA is dependent on the specific sequence to which it binds, elucidation of the mechanism responsible for the difference is likely to suggest how an apparently simple binding motif can participate in complex regulated gene expression. Our recent cloning of complementary DNA for GF-1 (manuscript in preparation) will facilitate more direct study of its expression, interactions with regulatory elements, and role in erythroid cell development. □

Received 26 October 1988; accepted 15 February 1989.

1. Stamatoyannopoulos, G. & Nienhuis, A. W. *The Molecular Basis of Blood Diseases* (ed. Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P. & Majer, P. W.) 66-105 (Saunders, Philadelphia, 1987).
2. Rutherford, T. R., Clegg, J. B. & Weatherall, D. J. *Nature* **280**, 164-165 (1979).
3. Bark, C., Weller, P., Zablinski, J., Janson, L. & Pettersson, U. *Nature* **328**, 356-358 (1987).
4. Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51**, 773-781 (1987).
5. Rosales, R. *et al.* *EMBO J.* **6**, 3015-3025 (1987).
6. Scheideit, C., Heguy, A. & Roeder, R. G. *Cell* **51**, 783-793 (1987).
7. Staudt, L. M. *et al.* *Nature* **323**, 640-643 (1986).
8. Wirth, T., Staudt, L. & Baltimore, D. *Nature* **329**, 174-178 (1987).
9. Ottolenghi, S. *et al.* *Blood* **71**, 815-817 (1988).
10. Surrey, S., Delgrosso, K., Malladi, P. & Schwartz, E. *Blood* **71**, 807-810 (1988).
11. Evans, T., Reitman, M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5976-5980 (1988).
12. Lingrel, J. B., Weimer, J. & Menon, A. in *Developmental Control of Globin Gene Expression* (eds Stamatoyannopoulos, G. & Nienhuis, A. W.) **251**, 201-210 (1987).
13. Mantovani, R. *et al.* *Nucleic Acids Res.* **15**, 9349-9364 (1987).
14. Lin, H. J., Anagnou, N. P., Rutherford, T. R., Shimada, T. & Nienhuis, A. W. *J. clin. Invest.* **80**, 374-380 (1987).
15. Jones, K. A., Yamamoto, K. R. & Tjian, R. *Cell* **42**, 559-572 (1985).
16. Siebenlist, U. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 122-126 (1980).
17. Mantovani, R. *et al.* *Nucleic Acids Res.* **16**, 7783-7797 (1988).
18. Superti-Furga, G., Barberis, A., Schaffner, G. & Busslinger, M. *EMBO J.* **10**, 3099-3107 (1988).
19. Shen, S., Slightom, J. L. & Smithies, O. *Cell* **26**, 191-203 (1981).
20. Seiden, R. F., Burke-Howie, K., Rowe, M. E., Goodman, H. M. & Moore, D. D. *Molec. cell Biol.* **6**, 3173-3179 (1986).
21. Tucker, K. A., Lilly, M. A., Heck, L. & Rado, T. A. *Blood* **70**, 372-378 (1987).
22. Rutherford, T. & Nienhuis, A. W. *Molec. cell Biol.* **7**, 398-402 (1987).
23. Gumuchio, D. L. *et al.* *Molec. cell Biol.* **8**, 5310-5322 (1988).
24. Collins, F. S., Stoeckert, C., Serjeant, G., Forget, B. & Weissman, S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4894-4898 (1984).
25. Ottolenghi, S. *et al.* *Blood* **69**, 1058-1067 (1987).
26. Wall, L., deBoer, E. & Grosfeld, F. *Genes Dev.* **2**, 1089-1100 (1988).
27. Dignam, J. D., Martin, P. L., Shastri, B. S. & Roeder, R. G. *Meth. Enzym.* **101**, 587-597 (1983).
28. Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Wiley, New York, 1987).
29. Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6523 (1981).
30. Strauss, F. & Varshovsky, A. *Cell* **37**, 889-901 (1984).
31. Raymondjean, M., Cereghini, S. & Yaniv, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 757-761 (1988).
32. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

ACKNOWLEDGEMENTS. We thank Sabra Goff, Seth Ruffins and Shawn Burgess for assistance, Bob Wise and Jim Kadonaga for technical advice, members of the laboratory for discussions, and Marie Fennell and Rebecca Hughes for preparation of the manuscript. We also thank Lise Riviere and Paul Tempst of the HHMI Core facility for oligonucleotides and the MIT Cell Center for K562 cells. Details on the experimental procedures used for transient expression experiments are available on request. This work was partially supported by the NIH. S.H.O. is an Investigator of the Howard Hughes Medical Institute.

Identification of the long ubiquitin extension as ribosomal protein S27a

Kent L. Redman* & Martin Rechsteiner

Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA

Two proteins of unknown function are encoded by 3' in-frame extensions of ubiquitin genes¹. The polypeptides are synthesized as an additional 52 (refs 2-4) or 76-80 (refs 4, 5) amino acids on the C terminus of ubiquitin, an unusual arrangement conserved in man^{3,5}, yeast⁴ and plants (J. Callis and R. Vierstra, personal communication). Although not homologous to each other or to ubiquitin, both extension proteins are highly basic and contain patterns of cysteine and histidine similar to those proposed to form 'zinc fingers'^{6,7}. The longer C-terminal extension protein (CEP80) is 30% lysine and arginine and, when denatured, behaves like a small cationic protein⁸. Its properties after isolation in physiological conditions, however, suggested that CEP80 is part of an RNA-protein complex. Using the antibodies that confirmed the presence of CEP80 in eukaryotic cells⁸, we show here that the protein is located on ribosomes. Immunoblotting of rat 40S subunit proteins specifically identifies CEP80 as ribosomal protein S27a.

Native CEP80 binds to DEAE resins unless pretreated with RNase and sediments at 100,000g (K.R., unpublished results). Analysis of a rabbit reticulocyte particulate preparation on sucrose gradients showed that CEP80 was present in fractions containing 80S monosomes and polysomes (Fig. 1a). Dissociation of ribosomal subunits with salt before sedimentation shifted CEP80 to fractions containing the 40S ribosomal subunit (Fig. 1b). In the light of these results, we sought to determine whether CEP80 is one of about 30 proteins known to be components of the eukaryotic small ribosomal subunit.

Like CEP80, most ribosomal proteins are basic, but the long extension is characterized by a lysine/arginine ratio of 3. Inspection of the amino-acid compositions for 33 rat^{9,10} and 24 *Artemia* small subunit proteins¹¹ revealed only four proteins with lysine/arginine ratios of more than 2.5 (Rat S3b, S12 and S27; *Artemia* S27a). S12 from rat is an acidic protein that does not correspond to the extension sequence¹²; S3b is twice the size of CEP80, as judged by SDS-PAGE¹⁰. Both rat S27 and *Artemia* S27a however, seem similar to CEP80 on the basis of their size and amino-acid composition, as analysed by the method of Cornish-Bowden¹³. Two-dimensional electrophoresis was used to determine whether either of these known ribosomal proteins is the long extension.

S27 and S27a are well separated from other small subunit proteins in the gel system used to name ribosomal proteins¹⁴, but the protein pattern in these acid-urea gels could only be partially transferred to nitrocellulose. A provisional identification of rat CEP80 was nevertheless possible as the protein detected with CEP80-specific serum was located between and below proteins S28 and S27 (data not shown). This position is occupied by only one known ribosomal protein, S27a^{14,15}. Rat ribosomal proteins have been examined in other two-dimensional electrophoresis systems, two of which use SDS-PAGE in the second dimension¹⁵. These protocols allow the reproducible transfer of proteins to nitrocellulose and, as shown in Fig. 2, our separations of rat 40S subunit proteins are comparable to published patterns¹⁵. Identification of CEP80 as 40S subunit protein 27a is based on the position of the immunodetected extension as well as its mobility relative to protein S27 in the two electrophoresis systems. S27a is the only small 40S

* Present address: Department of Biology, Molecular and Cellular Biology Section, University of Alabama, Box 870344, Tuscaloosa, Alabama 35487-0344, USA.

subunit protein that has significant shift in relative mobility between the two first-dimension gel systems¹⁵. CEP80, like S27a, migrates farther than S27 in the low pH system, but not as far as S27 in high pH gels (Fig. 2). These properties identify the long ubiquitin extension as ribosomal protein S27a.

Comparisons of amino-acid contents suggested that rat S27 is the long extension protein, but two-dimensional electrophoresis of 40S subunit proteins in three gel systems identified rat S27a as CEP80. Given these results and the similarity in amino-acid composition of rat S27 and *Artemia* S27a, we conclude that rat S27a was previously misidentified. Chemical modification during isolation may have altered the electrophoretic mobility of the purified rat protein⁹, whereas the isolation of *Artemia* S27a directly from electrophoretic gels would avoid this problem¹¹.

Identification of CEP80 as a previously recognized ribosomal protein does not prove that it has a ribosomal function. Co-isolation of the long extension protein with a single ribosomal

subunit reduces the possibility that the association is artefactual¹⁶, but subunit-specific adventitious binding has been observed previously¹⁷. CEP80 remains bound to the 40S subunit in salt solutions significantly above the 0.5 M concentration generally considered diagnostic for specific binding. Additional evidence for specificity was obtained by ion-exchange chromatography. Rabbit reticulocyte CEP80 elutes from DEAE columns with proteins that exactly match those of the 40S ribosomal subunit demonstrating that its association with the small subunit is independent of the isolation method (K.R., unpublished results). Further evidence for interaction between the long extension and ribosomes comes from *in vivo* studies of yeast strains lacking the gene for the yeast homologue of CEP80. These yeast grow slowly¹⁸, have a reduced ratio of small to large ribosomal subunits and inefficiently process the precursor for 18S ribosomal RNA¹⁹. We expect that a yeast small subunit protein will soon be identified as the 76 amino-acid yeast long-extension protein. Indeed, the deduced amino-acid

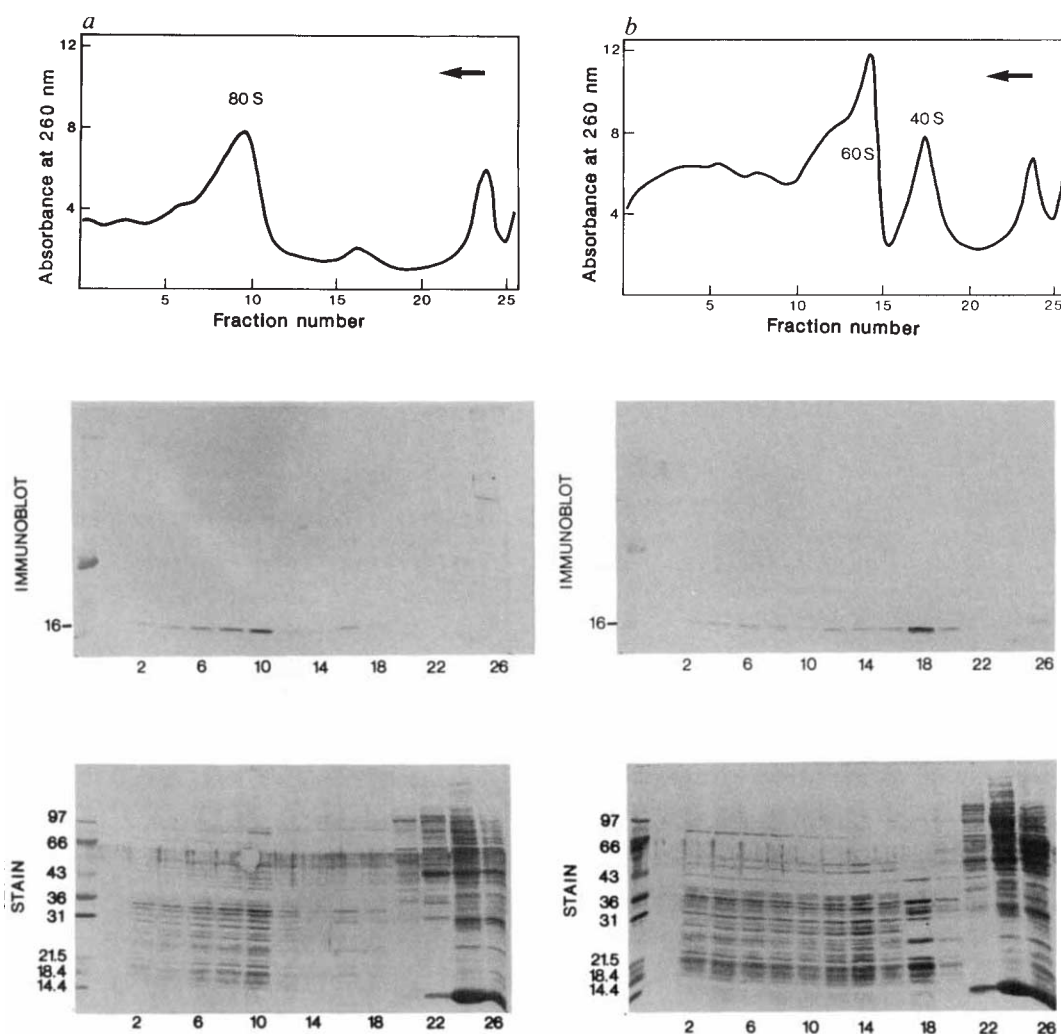


FIG. 1 Sedimentation of rabbit reticulocyte CEP80 in sucrose gradients. Resuspended polysomes were sedimented in low ionic strength buffer (a), or in high salt buffer to dissociate ribosomal subunits (b). Upper panels are profiles of absorption at 260 nm of gradients sedimented in the direction of the arrow. Middle panels are immunostained transfers of proteins from even-numbered fractions after separation by SDS-PAGE. Lower panels are transfers of identical gels stained with amido black. Relative molecular masses were estimated from prestained Rainbow markers (Amersham) on immunoblots and conventional markers⁸ on stained blots and are shown on the left, in thousands.

METHODS. Reticulocyte lysate prepared from rabbits²⁴ was centrifuged for

2 h at 100,000g. Samples of the pellet, containing 40–60 A_{260} units in 0.5 ml, were layered on to 11.5-ml linear gradients of 10–35% sucrose and centrifuged at 187,000g for 2.5 h. Aliquots of the 0.5 ml fractions were analysed by SDS-PAGE on 10–20% acrylamide gradient gels and electroblotted on to nitrocellulose⁸. Immunostaining (using extension protein-specific antiserum P7, goat-anti-rabbit peroxidase-coupled IgG and 4-chloro-1-naphthol) and protein staining were as previously described⁸. The sample and gradient in a were prepared in 50 mM Tris pH 7.8, 12.5 mM $MgCl_2$, 80 mM KCl, whereas the sample in b was prepared in the same buffer containing 880 mM KCl and analysed on a gradient containing 500 mM KCl.

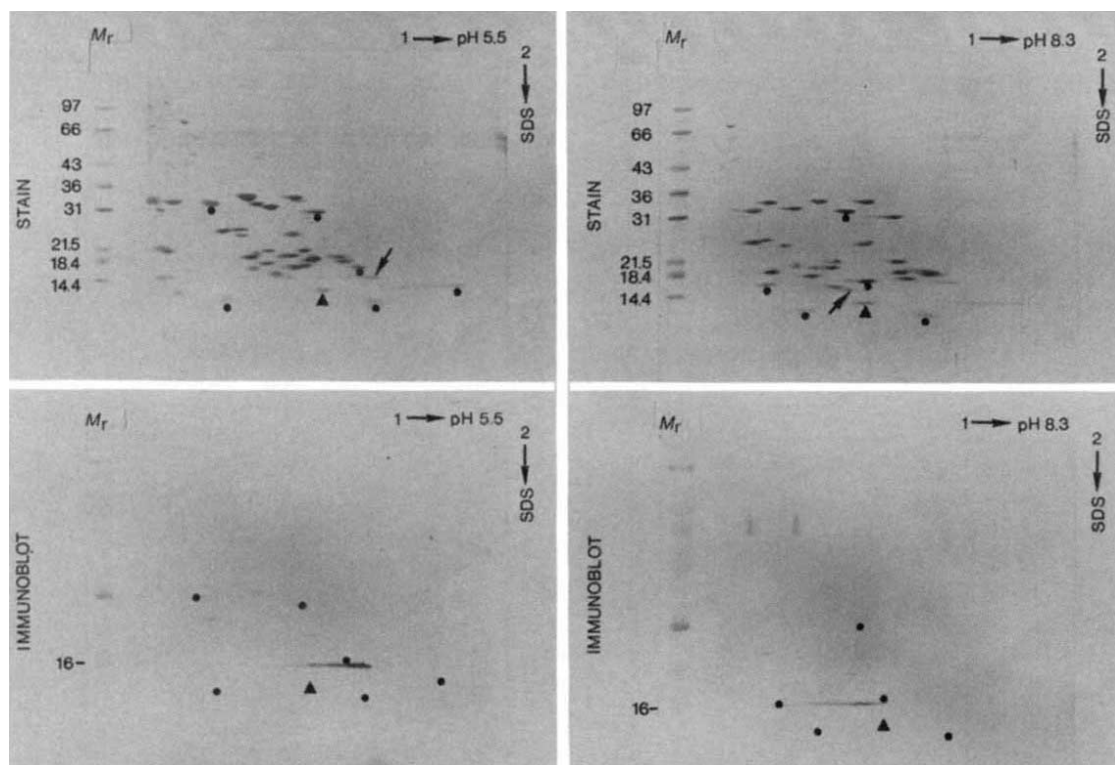


FIG. 2 Immuno-identification of the long ubiquitin extension. Identical nitrocellulose transfers of two-dimensional gels were amido-black stained or immunoblotted after reversible Ponceau S staining²⁵. Arrows on the stained transfers indicate the position of the immunoreactive protein. The position of S27 is indicated by a triangle. The relative molecular mass (M_r) markers used in Fig. 1 were added to the second-dimension gels (sizes in thousands). METHODS. 40S ribosomal proteins were prepared from the livers of Long-Evans rats as described¹⁵, and 80 μ g were separated on acid-urea/SDS

gels (left) or pH 8.3 Tris-borate/SDS gels (right)²⁶. SDS-PAGE was performed as described⁸, except that the concentration of Tris in the separating gel was increased to 0.56 M to improve the resolution of small proteins²⁷. Transfers were either stained with amido black (upper panels) or Ponceau S to locate individual proteins. After marking the positions of specific proteins on both blots (black dots), the Ponceau S transfers were de-stained and used for the immunodetection of CEP80 as in Fig. 1.

composition of the yeast long extension⁴ is nearly identical to that determined for yeast small-subunit protein YS 24 (ref. 20), a protein that is similar in mobility to mammalian CEP80 upon acid-urea/SDS two-dimensional gel electrophoresis²¹.

Why do eukaryotic cells synthesize a ribosomal protein fused to ubiquitin? This novel gene arrangement may have a regulatory role as the translation product is cleaved⁸, and synthesis of the extension protein as a ubiquitin fusion product is not required for wild-type growth in yeast¹⁹. Degradation of misprocessed, misfolded or otherwise short-lived polypeptides is an important function of ubiquitin-mediated proteolysis²². We propose that the fusion gene provides a mechanism to maintain a fixed ratio between ubiquitin, necessary for degrading aberrant proteins, and ribosomes, a major source of aberrant proteins. By binding to ribosomes and its own mRNA, S27a could regulate its own translation and that of ubiquitin. This form of translational regulation has been shown to occur for *Escherichia coli* ribosomal protein operons²³, but not for the monocistronic genes that encode eukaryotic ribosomal proteins. Although it remains to be seen whether our proposal is correct, conservation of this unusual gene arrangement among eukaryotes implies that there is an important physiological basis for co-synthesis of ubiquitin and a ribosomal protein. □

6. Berg, J. M. *Science* **232**, 485-487 (1986).
7. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1-3 (1988).
8. Redman, K. L. & Rechsteiner, M. *J. Biol. Chem.* **263**, 4926-4931 (1988).
9. Collatz, E., Wool, I. G., Lin, A. & Stöfler, G. *J. Biol. Chem.* **251**, 4666-4672 (1976).
10. Collatz, E. *et al.* *J. Biol. Chem.* **252**, 9071-9080 (1977).
11. Odani, S., Kenmochi, N. & Ogata, K. *J. Biochemistry* **103**, 872-877 (1988).
12. Lin, A., Chan, Y.-L., Jones, R. & Wool, I. G. *J. Biol. Chem.* **262**, 14343-14351 (1987).
13. Cornish-Bowden, A. *Analyt. Biochem.* **105**, 233-238 (1980).
14. McConkey, E. H. *et al.* *Molec. gen. Genet.* **169**, 1-6 (1979).
15. Madjar, J.-J., Arpin, M., Buisson, M. & Reboud, J.-P. *Molec. gen. Genet.* **171**, 121-134 (1979).
16. Hardy, S. J. S. & Kurland, C. G. *Biochemistry* **5**, 3676-3684 (1966).
17. Nue, H. C. & Heppel, L. A. *Proc. natn. Acad. Sci. U.S.A.* **51**, 1267-1274 (1964).
18. Finley, D. *et al.* in *Ubiquitin* (ed M. Rechsteiner) 39-75 (Plenum, New York), (1988).
19. Finley, D., Bartel, B. & Varshavsky, A. *Nature* (in the press).
20. Otaka, E., Higo, K. & Osawa, S. *Biochemistry* **21**, 4545-4550 (1982).
21. Otaka, E. & Osawa, S. *Molec. gen. Genet.* **181**, 176-182 (1981).
22. Rechsteiner, M. *A. Rev. Cell Biol.* **3**, 1-30 (1987).
23. Nomura, M., Gourse, R. & Baughman, G. A. *Rev. Biochem.* **53**, 75-117 (1984).
24. Brown, G. E., Kolb, A. J. & Stanley, W. M. *Meth. Enzym.* **30**, 368-387 (1974).
25. Salinovich, O. & Montelaro, R. C. *Analyt. Biochem.* **156**, 341-347 (1986).
26. Madjar, J.-J., Michel, S., Cozzone, A. J. & Reboud, J.-P. *Analyt. Biochem.* **92**, 174-182 (1979).
27. Fling, S. P. & Gregerson, D. S. *Analyt. Biochem.* **155**, 83-88 (1986).

ACKNOWLEDGEMENTS. This study was supported by the NIH. We thank Joyce Eshleman and June Reese for word-processing.

NATURE BACK ISSUES

SINGLE copies of Nature are available from the addresses shown below.

Prices: UK, £2.50; USA & Canada, US\$6.00 (surface), \$9.00 (air); Rest of World, £3.00 (surface), £4.00 (air).

UK & Canada
Nature,
65 Bleeker Street,
New York,
NY 10012, USA

UK & Europe
Marketing Department,
Nature,
4 Little Essex Street,
London WC2R 3LF, UK

Received 8 December 1988; accepted 22 February 1989.

1. Schlesinger, M. J. & Bond, U. *Oxf. Surv. Euk. Genes* **4**, 77 (1987).
2. Müller-Taubenberger, A., Westphal, M., Jaeger, E., Noegel, A. & Gerisch, G. *FEBS Lett.* **229**, 273-278 (1988).
3. Salvensen, G., Lloyd, C. & Farley, D. *Nucleic Acids Res.* **15**, 5485 (1987).
4. Özkaynak, E., Finley, D., Solomon, M. J. & Varshavsky, A. *EMBO J.* **6**, 1429-1439 (1987).
5. Lund, P. K. *et al.* *J. Biol. Chem.* **260**, 7609-7613 (1985).

Measuring the new superconductors

J. D. Doss

Faster, more straightforward methods are needed for testing the new high-temperature copper-oxide superconductors. A non-contact eddy-current characterization instrument particularly suitable for screening large numbers of samples will soon be on the market.

SEVENTY-EIGHT years ago in Leiden, the Netherlands, a small team headed by Heike Kamerlingh Onnes discovered superconductivity. The historic measurement on a column of frozen mercury was performed by Onnes's graduate student, Gilles Holst, who nulled a Wheatstone bridge in an attempt to measure a resistance that vanished below 4.2 K. Ever since that landmark measurement, the determination of d.c. resistance has remained the most common method for the characterization of superconductors. The Wheatstone bridge, of course, has now been replaced by sophisticated constant-current sources and microvoltmeters.

Even though d.c. resistance measurements are useful, the typical experimentalist requires additional measurement capabilities for superconductor characterization. Because zero voltage loss occurs only in static fields, other techniques are required to measure the small but finite losses found in the presence of alternating fields. Non-contact techniques are generally preferred (particularly in the study of superconducting thin films) because contacts tend to modify the superconductor surface.

Magnetic phenomena

With the information available to them, it was quite natural that Onnes and his contemporaries should believe that a vanishing d.c. resistance was the most fundamental indicator of superconducting behaviour. (It is not.) The most definitive property of superconductivity was not discovered until 1933, when Walter Meissner and Robert Ochensfeld demonstrated that magnetic flux is expelled from a superconductor as it is cooled through its transition temperature. The discovery of near-perfect diamagnetism ($\mathbf{B} = 0$) was the first clue that superconductivity is fundamentally a magnetic phenomenon¹.

The significance of the Meissner-Ochensfeld experiment is not lost on those who are investigating the new family of copper-oxide superconductors. While candidate superconductors may be identified initially through a d.c. resistance measurement, true superconducting behaviour is proven only by demonstration of the Meissner effect. The measurement of magnetization ($\mathbf{M}$) is also a routine procedure in the evaluation of superconductors. Much about the quality of a superconductor can be inferred from its

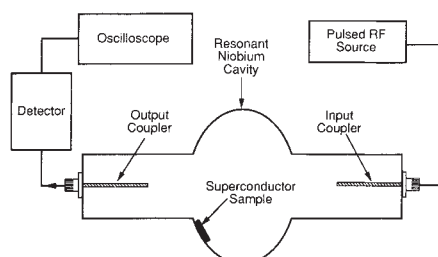


FIG. 1 The surface resistance of a superconducting sample may be determined by measuring its perturbing effect on the Q of a resonant cavity. The "decrement" method, used for measuring the high values of Q exhibited by superconducting cavities, is illustrated here.

susceptance (χ); indeed, the superconducting proportion of a multi-phase compound can be determined by measuring the flux expelled from a known volume of material. (Note: $\chi = M/H$; $\mathbf{B} = \mu_0[\mathbf{H} + \mathbf{M}]$ in SI units, while $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$ in gaussian units.)

Despite the value of susceptometry, the measurement of flux expulsion is not the most convenient method for screening large numbers of samples, primarily due to cost and time constraints. The magnetic susceptance of very small samples, such as films whose thickness is less than the London penetration depth (λ_L), is also very difficult to measure, because they exhibit no significant expulsion of magnetic fields. Thin films are of considerable interest for electronic applications of the new high-temperature superconductors, but other techniques must be used to determine their quality.

Surface resistance

Determining superconductor energy loss at frequencies ranging up to tens of GHz is important not only for classical radio frequency (r.f.) applications, but also for predicting the performance of superconductors in high-speed computer circuitry. A variety of techniques are employed to characterize superconductor loss at high frequencies, including several approaches where the sample is placed in some type of transmission line: subsequent perturbations are related to the sample's dissipation of energy².

A very effective method for determining the microwave surface resistance (R_s) of high-transition temperature (high- T_c) superconductors involves measuring the perturbation caused by placing the sample inside a resonant cavity. The cavity is first

calibrated by measuring the quality factor (Q) of the empty cavity, which is inversely proportional to the resistance of its inner surface, and the Q of the "loaded" cavity using a sample of similar geometry with a known R_s , such as stainless steel. With this data, one may determine the surface resistance of a superconducting sample measured in the cavity^{3,4}. For example, 3 GHz cavity perturbation measurements on bulk $\text{Ti}_2\text{Ca}_2\text{Ba}_2\text{Cu}_2\text{O}_{10}$, $\text{Bi}_2\text{CaSr}_2\text{Cu}_2\text{O}_8$, and $\text{YBa}_2\text{Cu}_3\text{O}_7$ yield R_s values ranging from $10^{-4}\Omega$ to $4 \times 10^{-3}\Omega$, compared to $\sim 10^{-6}\Omega$ for Nb (ref. 4).

Cryogenic copper cavities are useful in some instances, but superconducting cavities are generally preferred when higher sensitivity is required. The Q of non-superconducting cavities may be determined by standard bandwidth measurements, because these have relatively large bandwidths. (Q is equal to f_0/bw , where f_0 is the resonant frequency, and bw is the bandwidth within the -3 dB points.) When superconducting cavities are used, the direct determination of bandwidth is not always straightforward, because the Q of the empty cavity may exceed 10^9 . At a resonant frequency of 1 GHz, this implies a bandwidth of only 1 Hz, hardly a value one would wish to measure with a sweep-frequency generator. The higher Q s are measured by the "decrement" method to determine the fill (or decay) time (τ) of the cavity, which is driven from a pulsed r.f. source (Fig. 1). ($Q = \pi f \tau = \omega \tau / 2$, where f is the microwave frequency (usually near f_0) used for the measurement⁵.) While resonant-cavity measurements are the preferred method for determining R_s , simpler non-contact methods may be used to screen large numbers of samples to identify those specimens which are worthy of more detailed studies.

Eddy-current techniques

Precision "eddy-current" techniques have been used to determine the resistivity, transition temperature (T_c), and penetration depth (λ_L) of metallic and superconducting samples⁶⁻⁸. A modified eddy-current technique may also be used to characterize resistance losses in a qualitative manner, to determine T_c and the detailed structure of the superconducting phase transition⁹. As illustrated in Fig. 2, coupled inductors L_a and L_b may be used to induce r.f. currents in a superconductor

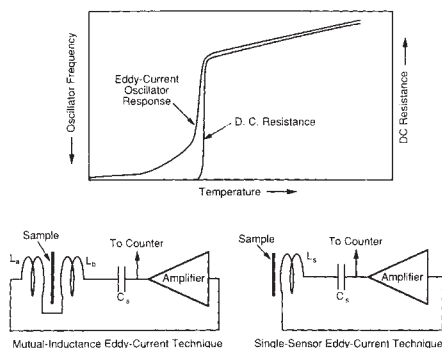


FIG. 2 Eddy-current analysis reveals energy losses below the critical transition temperature usually not apparent from d.c. resistance measurements. In d.c. measurements, a single thread-like superconducting path can "short" the contacts, effectively masking information about behaviour at lower temperatures. Two types of eddy-current oscillator circuits are shown.

sample. The induced current, a function of the sample's energy dissipation modifies not only the mutual inductance, but also the self-inductance of each coil. The inductive sensors are combined with capacitance (C_s) to form a series-resonant circuit. The addition of a wide-band amplifier to close the loop yields an oscillator whose frequency is a function of sample energy dissipation (losses).

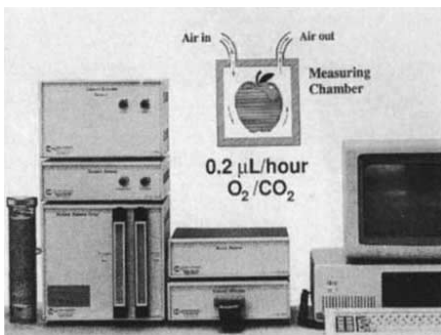
While the dual-coil system is less sensitive to microphonic noise, very effective systems can also be designed using a single sensor coil. In either case, oscillator frequency is plotted as a function of sample temperature to reveal details about energy losses, yielding information not available from d.c. resistance measurements at temperatures below T_c (ref. 10). While it is rarely a serious problem, eddy-current characterization does have a limited dynamic range. For extremely high or low values of sample resistivity, changes in resistance may not be sensed for a particular coil-sample configuration. Within limits, the dynamic range can be shifted to a more appropriate region by modifying the coupling between the sample and the sensor coil(s). Los Alamos National Laboratory has transferred its eddy-current technology to Lake Shore Cryotronics, Westerville, Ohio, which is preparing both single-sample and multiple-sample superconductor screening products for the market. □

James D. Doss is at the Medium Energy Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. For more information, fill in reader service number 100.

Quantitative aids

A top-loading microbalance, a micro-respirometer, a gaussmeter, and a range of pre-calibrated electrodes are a few of the latest measuring devices for the laboratory.

THE Micro-Oxymax **computerized metabolic computer** from Columbus Instruments measures oxygen consumption and CO_2 product simultaneously in 10 chambers, varying in size from 100 ml to 10



Micro-Oxymax: the newest respirometer from Columbus Instruments.

litres (Reader Service No. 101). Columbus Instruments says the Micro-Oxymax can be used to measure the metabolism of insects, the proliferation of bacteria or moulds, and the oxygen uptake of tissue cultures and fermenters, with a sensitivity down to $0.2 \mu\text{L}$. Measurements can be performed in ambient temperature or in a temperature-controlled waterbath. The measurement intervals can be specified between 5 minutes and 24 hours. The Micro-Oxymax measures the volumes of O_2 and CO_2 consumed or produced by the sample under study, and monitors the temperature of the sample chamber. The system is controlled by an IBM-PC or compatible computer; data is saved and reports are generated automatically.

SLM Instruments, Inc. introduced a new **cation measuring spectrofluorometer** earlier this month for use in studies involving intracellular calcium probes such as Fura-2 and Indo-1, and other multiple-wavelength fluorescent probes (Reader Service No. 102). The company's computerized model DMX-1000 Multiparameter Cation Measuring Spectrofluorometer is a combination of a high-speed scanning spectrofluorometer with a light-chopped excitation system. The spectrofluorometer's T-Optics configuration especially suits it for measuring multiple-excitation wavelengths, but it can also be

used for ratiometric measurements and polarization studies. Because of the modular nature of the model DMX-1000, polarizers, filters and other accessories can be installed as needed. Available accessories include a stopped-flow accessory, and a programmable cell changer that can handle four samples.

Genzyme Corporation has launched three new enzyme-based kits for **oligosaccharide analysis** (Reader Service No. 103). Genzyme's UnLinkit-N, AnaLink-N and EndoLink kits can be used to release asparagine-linked oligosaccharides from glycoproteins and glycopeptides, for subsequent analysis using HPLC. Genzyme's kits can be used with the Waters range of columns designed for analysing released oligosaccharides.

High T_c tech

The French company Drusch has a **nuclear magnetic resonance gaussmeter** for use in solid-state or nuclear physics laboratories (Reader Service No. 104).



The Drusch model RMN 2 nuclear magnetic resonance gaussmeter.

Drusch says the model RMN 2 gaussmeter is simple to operate: the user selects a probe appropriate for the expected field strength, places the probe in the field, and turns on the instrument. Probe alignment is not critical to the measurement, and the probe can be located up to 160 feet away from the readout. The RMN 2 is based on the measurement of the gyromagnetic ratio of protons as the input frequency is swept through the range of interest. Drusch says its gaussmeter can measure any field between 20 mT and 9 T with a precision of 5×10^{-3} mT, and a relative precision of at least 5×10^{-4} mT. The model RMN 2 can also automatically track a varying magnetic field at speeds of up to 10 mT per second.

Mettler's new modular model TA4000 **thermal analysis system** has an extended differential scanning calorimetry range of

1. Doss, J.D. *Engineer's Guide to High-Temperature Superconductivity* (Wiley, New York, in the press).
2. Sridhar, S. *Microwave J.* **30** no. 6, 117–123 (1987).
3. Sridhar, S. & Kennedy, W.L. *Rev. sci. Instrum.* **59**, 531–536 (1988).
4. Cooke, D.W. *et al. Phys. Rev. B* (submitted).
5. Ginzton, E.L. *Microwave Measurements*, 428–434 (McGraw Hill, New York, 1967).
6. Schawlow, A.L. & Devlin, G.E. *Phys. Rev.* **113**, 120–126 (1959).

7. Dalrymple, B.J. & Prober, D.E. *J. Low Temp. Phys.* **56**, 545–574 (1984).
8. Fiory, A.T., Hebard, A.F., Mankiewicz, P.M. & Howard, R.E. *Appl. Phys. Lett.* **52**, 2165–2167 (1988).
9. Doss, J.D., Cooke, D.W., McCabe, C.W. & Maez, M.A. *Rev. sci. Instrum.* **59**, 659–661 (1988).
10. Doss, J.D. *et al. Superconductor Science and Technology* (in the press).

-170 to +750 °C, and displays curves in real-time when connected to a PC (*Reader Service No. 105*). The system offers one-key automatic analysis for differential scanning calorimetry, thermomechanical analysis and thermogravimetric analysis. All test programs are permanently stored in the system processor, which can output a complete analysis record to a printer. The system's optional TA72 software allows the model TA4000 to work with an IBM AT or PS/2 computer for on-screen curve plotting, kinetic evaluations, and test data storage.

Balancing acts

The first **toploading microbalance** is new from Sartorius Instruments (*Reader Service No. 106*). Sartorius's model M3P is sensitive to 1 µg, and has three sensitivity ranges and a maximum capacity of 3 g. Unlike conventional microbalances,



Top-loading microbalance from Sartorius.

Sartorius's balance has no large weighing chamber with mechanical devices to extend and retract the weighing pan. Users can deposit and retrieve samples quickly, with no lockdown procedures to follow or doors to open and close. The microbalance calibrates automatically, and has built-in programs for data processing. The low-profile, compact model M3P has an RS232 interface for connection to a computer for saving or importing data to other programs.

The Japanese company A & D Company, Ltd has an **analytical balance** that comes with a wireless remote control and a vibrating spoon (*Reader Service No. 107*). The company's balance comes in two models: the FR-200 weighs samples up to 210 g in 0.1 mg increments, and the FR-300 handles up to 310 g with the same precision. The balance has a percentage and counting function, and can be used to weigh in grams, carats, pennyweights, decimal ounces, troy ounces, grain units, tolas, taels and mommes. A & D's analytical balance can be calibrated with the touch of a button, and has an interval data output feature that sends weighing data to a computer at set intervals. The balance

controls the vibrating spoon at frequencies between 110 Hz and 230 Hz.

The CP Instrument Company has a **pocket-size digital scale** that measures just 142 × 87 × 15 mm (*Reader Service No. 108*). The top-loading balance is ideal for use in the field, and saves space in the laboratory. Its weighing capabilities belie its small size: the scale measures up to



CP Instrument Co.'s balance fits in a pocket.

500 g in 1-g increments, and up to 1 kg with a precision of 2 g. The balance is described in CP Instrument Company's 160-page catalogue, which has ten pages devoted to balances.

Spectral studies

Minolta has a compact, **portable spectrophotometer** for colourimetric studies (*Reader Service No. 109*). The model CM-1000 has a flip-up LCD screen, much like a lap-top computer, and a hand-held spectral sensor which measures wavelengths between 400 and 700 nm using a 40-segment silicon diode array and a spectral filter array. The unit has a built-in printer, an RS-232 interface, and a 3.5-inch floppy disk drive. Statistical software allows spectral and colourimetric data to be displayed graphically, in two and three dimensions.

The model ELS-800 **electrophoretic light-scattering spectrophotometer** from Otsuka Electronics was developed for the determination of the surface charges of fine particles and electrophoretic polymers (*Reader Service No. 110*). The system can be used to study protein degeneration, the interaction between proteins and surfactants, the dispersion and aggregation of liposomes and vesicles, and the properties of surfactants and micelles. The model ELS-800 uses the optics of the Heterodyne dynamic scattering method to yield measurements of the zeta potential and electric mobility distribution based on the laser Doppler electrophoretic method. The system's correlator software employs a fast Fourier transform.

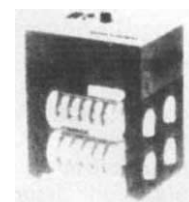
Image analysis

The BioVision personal **microscope workstation** from Perceptics puts the power of an Apple Macintosh II with a high-resolution colour monitor to work with an optical microscope, an image sensor, a frame grabber, and TCL-Image biological software designed for biomedical research (*Reader Service No. 111*). Perceptics designed the workstation for entry-level image analysis: the company says the mouse, icons, windows and pull-down menus of the Macintosh reduce learning time and simplify operation. The system's frame grabber provides real-time image capture with 256 gray levels. The workstation's TCL-Image software performs arithmetic, linear and non-linear neighbourhood, binary, and morphological operations, along with distance, image and Fourier transformations. Users can also append their own image processing functions, and develop application programs using the existing library functions.

Digithurst Ltd offers a range of **low-cost colour image analysis systems** at prices

ADVERTISEMENTS

DIABORM EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

DIABORM
Geräte
P.O. Box 850126
08000 Munich 85,
FRG

Reader Service No.49

Economical Modular Tissue Culture Incubator



- Safe for AIDS research • Reliable, for In-Vitro fertilization
- Versatile enough for thousands of lab uses

Easy to use - Just flush with gas mixture, seal and place at approp. temp. Perform multiple experiments simultaneously w/out cross contamination. Thousands in use worldwide.

For information contact: **Flow Laboratories, Ltd.** or **Billups-Rothenberg, Inc.** • P.O. Box 977
Del Mar, CA 92014-0977 USA • (619) 755-3309

Reader Service No.25

comparable to monochrome systems (*Reader Service No. 112*). Digithurst's microcomputer-based MicroScale IC for entry-level applications and MicroScale IIC for more advanced studies range in price from under £1,000 to £12,000 (UK). Both packages offer object measurement, including area, perimeter, position, diameter, orientation and volume, and whole-field measurement of area, perimeter, and average intensity. Because every pixel within an image is quantified in



Low-cost image analysis from Digithurst.

terms of its colour content, not just its intensity (as in a monochrome system), the MicroScale systems can cope with extremely complex colour combinations. Thresholds of colour are set manually with the MicroScale IC, and automatically or manually with the MicroScale IIC.

Electrochemical sensors

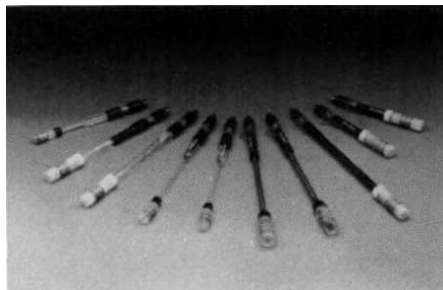
Diamond General's Chemical Microsensor II can be operated simultaneously in the amperometric, potentiometric and thermal sensing modes (*Reader Service No. 113*). The **ion electrode** can determine the concentrations of chemical species such as pH, CO, PO, and PH. The compact Chemical Microsensor II fits easily on



Measure most any ion with the Chemical Microsensor II.

the lab bench, and its battery pack allows it to be carried into the field for performing on-site measurements.

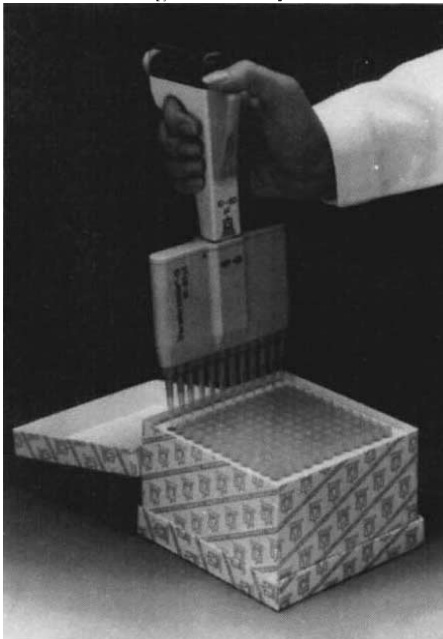
Beckman Instruments has six new **calomel combination electrodes** designed especially for use in biologicals and other silver-sensitive solutions, including Tris buffers, enzyme preparations, tissue culture media, sulfides, bromides and iodides (*Reader Service No. 114*). The Beckman FUTURA Plus electrodes are offered in six styles: with glass bodies measuring 12 mm × 130 mm, 10 mm × 195 mm, 5 mm



Beckman's array of pre-calibrated electrodes. × 225 mm, and 5 mm × 178 mm; and with epoxy bodies measuring 12 mm × 130 mm and 7 mm × 225 mm. The electrodes can be used at temperatures of between 23 and 140 °F. Each electrode comes preconditioned in its own Performance Pac vial, and can be used right out of the box, with no soaking.

Pipette products

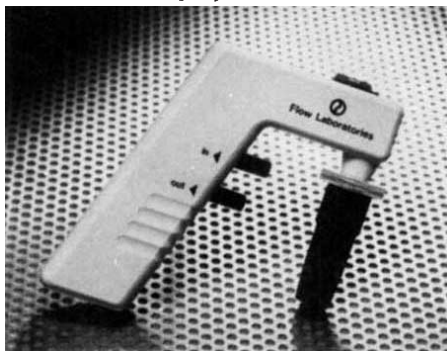
Rudolf Brand GmbH has two new **multi-channel pipettes** which can dispense eight or twelve volumes of liquid between 5 µl and 200 µl, at one shot (*Reader Service No. 115*). The Transferpette-8 and Transferpette-12 are ergonomically shaped for comfort, and the volume-setting, pipetting and tip-ejection functions are separated to avoid errors. Brand says the pipettes can be used to fill a large number of microtitre plates without giving the user a cramped hand, because the user's thumb always remains in a natural position. Both pipettes weigh roughly 150 g, and a digital display shows the set volume. The dosing portion of the pipettes can be screwed off for autoclaving. Brand says the Transfer-



Eight- or twelve-channel pipetting from Brand. pettes function with errors in accuracy of less than 1.5 per cent of the nominal volume.

Flow Laboratories recently launched a cordless, **rechargeable pipetting controller**

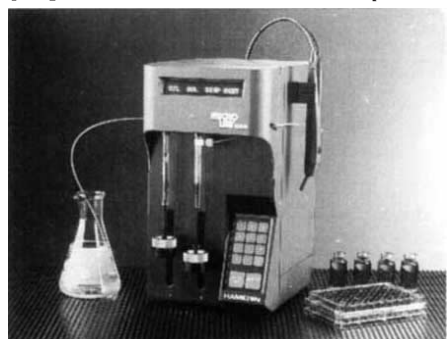
designed for use with pipettes ranging in size from 0.1 ml to 100 ml (*Reader Service No. 116*). Named the Pipetus Akku, the controller weighs 180 g, operates for up to five hours on one charge. When set at minimum power, operation can be extended to 15 hours without recharging. Flow's controller was designed for cell culture media preparation, and comes



Flow's cordless, rechargeable pipette controller.

complete with a recharging unit for £175 (UK). The Pipetus Akku features a 0.2-mm hydrophobic PTFE membrane filter with luerlock connections for air filtration. With normal use, Flow says the unit is maintenance-free, except for periodic filter replacement. Two safety features are included: a safety valve prevents flooding of liquid into the controller, and the entire controller can withstand fumigation with formaldehyde. Dispensing and filling are controlled by separate buttons.

Hamilton's Microlab 1000 **diluter/dispenser** has been designed to accommodate liquid volumes of between 1 µl and 25 ml (*Reader Service No. 117*). Users can program the Microlab 1000 with up to 50



The new programmable diluter/dispenser from Hamilton boasts 50 protocols.

protocols by answering a series of yes-or-no questions displayed on the instrument's LED. The diluter/dispenser then automatically calculates optimum syringe size, valve size and syringe speed for each application. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.